

JOURNAL ARTICLE

A replicable acoustic measure of lenition and the nature of variability in Gurindji stops

Thomas Ennever^{1,2}, Felicity Meakins^{1,2} and Erich R. Round^{1,2}

¹ University of Queensland, St Lucia, QLD 4072, AU

² ARC Centre of Excellence for the Dynamics of Language, AU

Corresponding author: Erich R. Round (e.round@uq.edu.au)

An automated method is presented for the commensurable, reproducible measurement of duration and lenition of segment types ranging from fully occluded stops to highly lenited variants, in acoustic data. The method is motivated with respect to the relationship between acoustic and articulatory phonetics and, through subsequent evaluation, is argued to correspond well to articulation. It is then applied to the phonemic stops of casual speech in Gurindji (Pama-Nyungan, Australia) to investigate the nature of their articulatory targets. The degree of stop lenition is found to vary widely. Contrary to expectations, no evidence is found of a positive effect on lenition due to word-medial (relative to word-initial) position, beyond that attributable to duration; nor do non-coronals lenite more than their apical counterparts, which freely lenite along a continuum towards taps. No significant effect is found of preceding or following vocalic environment. Taken together, the observed lenition, duration, and peak intensity velocities are argued to be inconsistent with a single, fully-occluded articulatory ‘stop’ target which is undershot at short durations, rather targets can be understood to span a range or ‘window’ of values in the sense of Keating (1990), from fully-occluded stop-like targets to more approximant-like targets. It is an open question to what degree the patterns found in Gurindji are language particular, or can be related to the organization of obstruent systems in Australian languages more broadly. Precisely comparable studies of additional languages will be especially valuable in addressing these questions and others, and are possible using the method we introduce.

Keywords: Lenition; stops; articulatory phonology; acoustic phonetics; intensity; Gurindji; Australian languages

1. Introduction

The phonemic obstruent systems of Australian languages are systems of contrasting extremes. In one dimension, they host an abundance of place of articulation contrasts, particularly in the coronal region, and these are increasingly well understood (Anderson & Maddieson, 1994; Bundgaard-Nielsen et al., 2012, 2015; Butcher, 1995; Proctor et al., 2010; Tabain & Butcher, 2015; Tabain & Rickard, 2007). In all other dimensions, they are impoverished: Most possess just a single obstruent series, with no contrast in laryngeal features, length, or between stops and fricatives (Busby, 1980; Evans, 1995). Nevertheless, allophonic stop lenition patterns are widely reported in descriptions of Australian languages, and raise the question of exactly how the parametric space of ‘manner of articulation’ is utilized within Australian languages. The investigation of such matters bears on theories that propose language-specific influences on gestural target setting (Keating, 1990; Guenther, 1995).

An open research question in Articulatory Phonology and Task Dynamic approaches is whether gestural targets are to be construed as single points (Saltzman & Munhall, 1989)

or as ‘windows’ or ‘ranges’ of targets (Keating, 1990). To this end, we are interested in whether the lenition of obstruents in one Australian language can be explained as i) the mechanical byproduct of temporal reduction causing undershoot relative to a single, point-like target, ii) due to other known factors effecting stop-lenition in a similar manner, or iii) due to speakers actively selecting among multiple available target articulations within a range or window.

In order to answer these questions using acoustic data we present a novel method for deterministically and automatically demarcating phonemic stops and their allophonic variants, and deriving quantitative measures of lenition using intensity data. Detailing this method, and assessing it, comprise a major contribution of the paper.

We then proceed to a fine-grained acoustic-phonetic study of the realizations of single-series phonemic obstruents in an Australian language with respect to manner of articulation and lenition, with particular attention to the synchronic phonetic variability of phonemic obstruents in casual speech. We investigate phonemic stops in Gurindji (Ngumpin-Yapa subgroup of Pama-Nyungan) and ask the following questions:

1. What is the range of realizations (in terms of lenition) of the phonemic stops in Gurindji, and their relative frequencies?
2. Are these influenced by a stop’s place of articulation, vocalic environment, and/or word boundary adjacency, and if so, how?
3. Is there evidence to support an analysis of Gurindji stop phonemes having a single, fully-occluded point-like articulatory target, with more lenited variants the product of undershoot due to short duration; or conversely, is there evidence for a window-like range of articulatory targets?

To answer these questions, we study intervocalic realizations of four Gurindji phonemic stops /p t ʈ k/ in the casual speech of a female speaker. The paper is organized as follows. Section 1 provides a background to Australian obstruents, common patterns of allophony, and establishes known factors affecting stop lenition. We also survey the challenges posed by gradient phonetic variation and the need for robust techniques for the analysis of casual speech. Section 2 introduces the materials used in the study. In Section 3 we introduce and evaluate an automated procedure for delimiting, in a commensurable manner, stop-like and approximant-like segments from acoustic, casual speech data and estimating their degree of lenition. This research tool is applied to the Gurindji data in Section 4. Results for factors affecting lenition are presented in Section 5. Implications for the types of articulatory targets underlying phonemic stops in Gurindji are discussed in Section 6. Section 7 concludes.

1.1. Gurindji

Gurindji is a Ngumpin-Yapa (Pama-Nyungan) language spoken in the Victoria River District of the Northern Territory, Australia. It is the traditional language of the Gurindji people who live in the communities of Kalkaringi and Daguragu (Meakins et al., 2013). It is currently endangered with approximately 40 speakers remaining. Younger generations now speak the mixed language Gurindji Kriol (McConvell & Meakins, 2005).

1.1.1. Phoneme inventory

Gurindji’s phonological inventory is typical of many Pama-Nyungan languages, comprising a five-way place of articulation distinction for obstruents and corresponding nasals, three laterals, three glides, and a tap/trill, shown in **Table 1**. Gurindji makes no contrasts in terms of voicing, consonant length, or frication, and accordingly obstruents are transcribed in **Table 1** using the conventional voiceless IPA symbols. Phonetically, the pre-palatal obstruent /c/ is realized consistently as an affricate by the speaker we

study (Ennever, 2014b) and so is excluded from the present study. Like many Australian languages, the vowel system of Gurindji is sparse, contrasting three qualities and length (Meakins et al., 2013), shown in **Table 2**.

1.1.2. Morphological and prosodic structure

Primary stress falls on the initial syllable of Gurindji words without exception. The stress system has not been studied in detail, though broadly speaking, it resembles those of many Pama-Nyungan languages, with secondary stress on most suffix-initial syllables and alternating stress otherwise (Dixon, 2002, p. 557). A consequence is that word-initial syllables fall at the left boundaries of both the morphosyntactic word and a prosodic word. In this study, we examine intervocalic stop phonemes in word initial position (i.e., flanked on the left by the final vowel of a preceding word), and in morpheme-medial position. These two positions contrast in terms of (non-)adjacency to both morphosyntactic and phonological word boundaries. Given the state of knowledge of Gurindji’s stress system, we make no specific claims about foot boundaries, other than to note that word-initial tokens will always be foot-initial also.

1.2. Phonemic obstruents in Australian languages

Australian languages are known for their rich place of articulation distinctions, particularly among coronals—languages contrast either one or two apical articulations, plus one or two laminal articulations (Busby, 1980). Gurindji follows the double-apical pattern, contrasting apical alveolar and apical retroflex articulations, in addition to a single laminal pre-palatal place and the non-coronals; a bilabial and a dorso-velar. Cross-linguistically in Australia, alveolar phonemes vary in their precise point of contact with the alveolar ridge and retroflexes vary in terms of posterior placement and sublaminal contact (Chadwick, 1975; McGregor, 1990; Tabain, 2009). Even in languages that contrast two apical places, the contrast is typically neutralized word initially (Butcher, 1995; Tabain & Butcher, 2015; Steriade, 2001). This is true also in Gurindji.

Australian languages are also known for their paucity of manner distinctions, particularly among obstruents (Butcher, 2006). Only a handful of Australian languages possess phonemically contrastive fricatives, or stops that contrast in phonation or length (Butcher, 2004; B. Evans & Merlan, 2004; Evans, 1995, p. 730; McKay, 1980; Stoakes et al., 2007). Gurindji is typical in this sense, lacking any laryngeal, length, or manner contrast among obstruents.

1.2.1. Synchronic allophony

Allophonically, stops in Australian languages are commonly reported to possess lenited variants when flanked by vowels and/or liquids (Dixon, 2002; Evans, 1995). Non-coronal

Table 1: Gurindji consonant phonemes after Meakins et al. (2013). Orthography is in parentheses.

	Bilabial	Alveolar	Retroflex	Pre-palatal	Velar
Stop	p (p)	t (t)	ɽ (rt)	c (j)	k (k)
Nasal	m (m)	n (n)	ɳ (rn)	ɲ (ny)	ŋ (ng)
Lateral		l (l)	ɭ (rl)	ʎ (ly)	
Tap/Trill		r (rr)			
Glide	w (w)		ɻ (r)	j (y)	

Table 2: Gurindji vowel inventory (Meakins et al., 2013). Orthography is in parentheses.

	Front	Central	Back
High	ɪ (i), ɪ: (ii)		ʊ (u), ʊ: (uu)
Low		e (a), e: (aa)	

and palatal stops may possess corresponding glide allophones, and alveolar stops flapped or tapped allophones. Fricative allophones are less common but have been reported in similar environments (Fletcher & Butcher, 2015; Dixon, 2002). In terms of positional factors, word initial lenition is generally dispreferred, although some Australian languages have lenited allophones in word-initial position (Blevins, 2001). Lenition has been correlated with stress in Murrinh Patha (Mansfield, 2015) and Yir Yoront (Alpher, 1988). Most reports of allophony are impressionistic; however, Ingram et al. (2008) investigate spectrographic data to identify a range of connected speech processes involving reduction in Warlpiri, a Ngumpin-Yapa language related to Gurindji. These include: Stop voicing, trilling, nasal weakening, vocalization, deletion, nasal-stop cluster reduction, and labialization. Other than Ingram et al. (2008), much of the instrumental phonetic work conducted on Australian languages has focused either on place of articulation (Bundgaard-Nielsen et al., 2012, 2015; Butcher, 1995; Tabain, 2009; Tabain & Butcher, 2015) or on those few languages that contrast two series of stops (Butcher, 2004; B. Evans & Merlan, 2004; McKay, 1980; Stoakes et al., 2007). Here we address the resulting gap in our understanding of Australian languages, with respect to manner of articulation.

1.3. Known potential factors in stop lenition

1.3.1. Duration

One of the most commonly cited factors affecting lenition is rate of speech and segmental duration (Donegan & Stampe, 1979; Gurevich, 2008; Lindblom, 1983, 1990; Shockey & Gibbon, 1993; Zwicky, 1972). Kirchner summarizes the relationship (2001, pp. 217–218):

“...fast speech, by definition, involves shortening of articulatory gestures. This shortening can mean one of two things: either the articulator reaches the target constriction faster, or the constriction itself is shorter.”

It is under these conditions that we also expect articulatory undershoot resulting in acoustic lenition. Soler and Romero (1999), for example, find duration and degree of constriction to be highly and positively correlated in Spanish spirantization phenomena. In the Scouse variety of English, Marotta and Barth (2005) find fricative and approximant allophones to be successively shorter than their stop counterparts. Furthermore, the relationship is understood to be gradient rather than categorical. In American English, stop lenition is reported to be increasingly frequent and pronounced at successively quicker speech rates, and in successively less formal registers (Warner & Tucker, 2011). Kirchner (2001, p. 4) proposes an implicational hierarchy to this effect, claiming that “if a consonant lenites in some context, at a given rate or register of speech, it also lenites in that context at all faster rates or more casual registers of speech.” Taken together, these studies would suggest that, *ceteris paribus*, the shorter the duration afforded to a constriction, the less likely full constriction will be achieved.

1.3.2. Place of articulation

Place of articulation of the target segment has also been suggested to affect lenition. Foley (1977), for example, proposes a strength hierarchy of places of articulation ordered by their likelihood of undergoing lenition: Velar > bilabial > alveolar. Evidence supporting this is generally constrained to studies of the Romance languages—for example Florentine Italian (Dalcher, 2006) and Balearic Catalan (Wheeler, 2005, pp. 320–324). Divergent patterns are reported in many of the world's languages (see Kaplan, 2010 for a summary). Explanations for differences in lenition rates based on place of articulation have been couched in terms of physiological and aerodynamic factors (see Lavoie, 2001, pp. 133–138 for velars; Hualde & Nadeu, 2011 for bilabials).

Within Australia, evidence for place of articulation effects is typically marshaled from the extensive reconstruction of diachronic sound changes. One of the most striking sound changes affecting a number of Australian languages is word initial weakening and the loss of stops consonants—a process affecting bilabial, laminal, and velar obstruents but to the exclusion of apicals (Blevins, 2001; Koch, 2004). An additional set of well established historical changes concern languages that formerly possessed a two-way stop contrast. In a subset of these cases we find that the obstruent system has reduced to a single stop series for all places of articulation except for the apicals where a stop contrast is maintained (as for example in some dialects of the Yolngu languages) (Wood, 1978). The accepted path for this phonological re-organization is an intermediate stage of stop-glide lenition affecting the lenis peripheral and laminal stops (Dixon, 2002). Similarly, in the synchronic domain, Mansfield's (2015) sociophonetic study of lenition in Murrinh Patha notes that peripheral stops are more prone to lenite to approximants than coronal stops. Finally, cross-linguistic surveys of morphophonological alternations similarly demonstrate that peripheral and pre-palatal obstruents undergo lenition more frequently than their apical counterparts (Round, 2010).

Nevertheless, there is also synchronic and diachronic evidence for apical lenition. Taps are found as allophones of apical stop phonemes in a number of languages (see Dixon, 2002) and have been implicated in an intermediate stage of stop allophony preceding the emergence of three rhotic phoneme systems in the Karnic languages *inter alia* (Breen, 1997; Dixon, 2002). Despite alternations between stops and taps seemingly constituting lenition (i.e., shorter and less complete constrictions), there are no studies closely examining the acoustic properties of taps in Australian languages. Outside of Australia it has been noted that realizations of intervocalic voiced stops, typically transcribed as 'taps,' may include some formant structure—a feature more commonly associated with approximants (as reported in American English [Warner & Tucker, 2011]). Since taps have only been impressionistically noted in Australian languages, it is possible that the degree of apical lenition has been understated.

1.3.3. Flanking vowel quality

The present study focuses on the realization of phonemic stops in intervocalic position, widely accepted as the segmental environment most favourable for consonantal lenition (Kirchner, 2001; Lass, 1984, p. 182).¹ There is, however, ongoing research into whether the *quality* of the flanking vowels themselves has a significant impact on lenition outcomes. Within effort-based models (e.g., Kirchner, 2001, 2004), vocalic openness (or height) is argued to influence lenition rates due to the greater tongue displacement required to make oral closure. Perceptual-based models (e.g., Kingston, 2008) instead contend that consonantal lenition is not sensitive to vocalic openness. Within a perceptual approach, speakers are understood to attend to disparities in intensity between an affected (lenited) segment and its neighbors. In this view, lenition is motivated by a constraint against abrupt interruptions to the intensity contour of a particular prosodic unit, such as those created by a fully occluded stop. Since the intensity differences between consonants and vowels are much larger than the intensity differences between individual vowel qualities, it is argued that consonantal openness is a significant factor in motivating lenition but vocalic openness is not.

Empirical evidence on this issue however is scarce and, as of yet, inconclusive. Competing evidence is found in studies of Spanish lenition alone: Simonet et al. (2012) find less constricted realizations of /d/ after lower vowels than after high vowels, while Colet

¹ Often this environment is extended to include 'inter-continuant' environments incorporating all flanking domains involving vowels, glides, or liquids (cf. Kirchner's [2001] 'quasi-intervocalic').

et al. (1999) and Ortega-Llebaria (2004) find more constricted realizations of /g/ between low vowels. Straightforward expectations arising from claims of articulatory effort are further complicated by the possibility of the consonant in question shifting its place of articulation to co-articulate with the flanking vowels—or vice versa (cf. Carrasco et al., 2012, p. 169). Saltzman and Munhall (1989) find that in cases where there are competing constraints on articulators between vowels and consonants (e.g., [g] in environments /aga/ and /igi/), the *location* but not *degree* of constriction for the consonant will vary as a function of the overlapping vowel. There are even fewer studies of Australian languages that have investigated effects of flanking vowel quality on lenition outcomes. Mansfield (2015) reports that following vowel quality was not statistically significant in his study of /p/ and /k/ lenition in the Australian language Murrinh Patha once lexical item was included as a random effect.

We therefore include preceding and following vocalic environments in the current study to probe if there are any significant differences in lenition outcomes on the basis of articulatory effort. We group the vowels based on their proximity to the target consonant's constriction location. This differs from studies that split vocalic environment into 'open' and 'non-open' vowels. Instead we anticipate some effort reduction and therefore *less* lenition for /p/ and /k/ in the environment of /u/ since the former involves lip rounding and the latter involves tongue backing, both of which are articulatory features shared with /u/. In the case of /t/ and /t̪/ we cautiously anticipate greater co-articulation (and less lenition) in the environment of /i/ due to tongue tip raising, in contrast with /a/ and /u/.

1.3.4. Domain position effects

One final relevant factor affecting lenition outcomes is the position of the target segment within relevant domains. Escure (1977, p. 58) proposes an implicational hierarchy of positions in which lenition operates. She observes that initial lenition is generally less frequent than non-initial lenition at the level of the syllable, word, and utterance. The proposed hierarchy claims that if a language exhibits lenition domain-initially, it will also exhibit lenition in all other non-initial environments. While Escure's implicational hierarchy has been shown to be violated by a number of languages (see Bauer, 2008), its basic proposal of a dispreference for domain initial lenition has been widely borne out by cross-linguistic surveys (cf. Ségéral & Scheer, 2008). One explanation advanced for this is the importance of preserving phonological information in word onsets, which have been shown to contain acoustic cues critical to word-perception (Marslen-Wilson & Zwitserlood, 1989).

It is also the case that position affects duration (Oller, 1973; Edwards et al., 1991; Tabain, 2003; Cho, 2006), which in turn affects lenition (Section 1.3.1). Consequently we will be interested in this study to probe whether the contributions to lenition of duration and position are to some degree independent.

Finally, usage-based models (e.g., Bybee, Pierrehumbert) predict that tokens in high frequency lexical items are more prone to lenite than tokens in low frequency lexical items.² Such a prediction has been borne out by several lenition studies (Bybee, 2002; Pierrehumbert, 2001; Dalcher, 2006) and so lexical item is included in the present study as a random effect.

1.3.5. The abstract representation of segments

The concrete articulation of a phonetic segment can be regarded as an execution of a more abstract motor plan and/or phonological representation. Theories like Articulatory

² Where 'high' and 'low' frequency may be variously defined as some percentage of most utilized lexical items in the corpus/corpora.

Phonology (Browman & Goldstein, 1989) propose that such plans contain articulatory targets that may or may not be physically reached given other constraints such as segment duration. Specifically, sequential gestural units can be subject to effects of ‘intergestural sliding’ (Saltzman & Munhall, 1989). That is, when speech rate increases, articulatory gestures tend to ‘slide into each other,’ increasing their temporal overlap, and resulting in the truncation of one or both adjacent gestures. Such processes are typically assumed to be governed by *point-attractor* dynamics: Articulatory trajectories for a given gestural unit converge on a single state over time, i.e., a single specified target (Saltzman & Munhall, 1989). If this were the case, we would expect any failure to reach the specified target to be the result of duress, such as applied by temporal reduction. On the other hand, if articulatory trajectories need not converge on a single, point-like gestural target but rather a window-like range, there would be grounds for speakers freely producing a range of articulatory velocities and constriction degrees, at least partially independent of temporal reduction.

Parrell (2011) examines Spanish /b/, which like Gurindji stop phonemes, has many unoccluded, sonorous phonetic realizations. Parrell argues that a single, fully occluded articulatory target is sufficient to account for the variation in Spanish /b/, with other realizations the result of articulatory undershoot due to short duration. Parrell also observes that if Spanish /b/ had only an unoccluded target, then one would not expect occluded variants, even under conditions of long duration, yet long, occluded stops are precisely what are found. Like Spanish /b/, the stop phonemes of Gurindji are sometimes fully occluded, thus we have no reason to believe they are represented or planned solely with unoccluded targets. However, will we ask the question, of whether a single, fully occluded target is sufficient to account for the Gurindji data, or whether it is more consistent with there being a range of targets (or the target itself being represented as a range rather than a point), which span full occlusion through to more open articulations.

To be able to answer these kinds of questions acoustically, it is necessary for studies to be able to quantify gradient acoustic variation (such as that involved in stop lenition) and query the extent to which, and the circumstances in which, speakers may diverge from a kinematic system that assumes a point-like articulatory target and set temporal constraints.

1.4. The need for robust techniques of acoustic, casual speech analysis

We aim to infer properties of lenition from acoustic, casual speech data. Ideally, one might study lenition using articulatory data collected under laboratory conditions, however in practice there are good reasons also to pursue alternatives. For many lesser-studied languages, acoustic recordings of casual speech already exist whereas controlled articulatory data is unlikely for logistical reasons to become available in the near future. For languages no longer spoken, acoustic recordings may be all we can ever access. It is reasonable also to expect that casual speech will contain informative variation that may not be apparent in controlled lab speech; as Ohala (1996, p. 206) observes, “[t]he more we look at connected speech in detail, the larger the ‘zoo’ of strange and exotic phonetic animals becomes.” To understand lenition synchronically and diachronically, we wish to be able to study as much of the ‘zoo’ as possible.

1.4.1. Challenges of acoustic speech segmentation

Notwithstanding the advantages just mentioned of acoustic, casual speech data, its analysis presents well-known challenges. The segmentation of continuous speech into discrete acoustic or phonetic units is a somewhat artificial task (Turk & Sugahara, 2006). Ladefoged (2003, p. 103) cautions that “many segments [simply] don’t have clear beginnings and ends” and Fry (1979, p.117) goes so far as to declare that “[from the acoustic

point of view] there are only sounds which are more like, and sounds which are less like the vowels of voiced speech.” Concretely, the segmentation of speech sounds presents three challenges: (i) Discretization, (ii) commensurability, and (iii) reproducibility. By ‘discretization,’ we mean the challenge of delineating the edges, by whatever means, of speech sounds. Many speech sounds, whether viewed acoustically or articulatorily, have no point-like onset and offset events, and consequently various proxies are resorted to (Fant, 1973; Lavoie, 2001). **Table 3** presents criteria employed for segmenting regular ‘oral stops’ in some recent studies of stop lenition.

By ‘commensurability’ we refer to the challenge of comparing across different segment types. For example, if one uses ‘bursts’ to define the right edge of a true phonetic stop, how should this be compared to the right edge of allophonic variants such as taps (Connell, 1991), fricated stops (Dalcher, 2006), or simple approximants? In Gurindji, this is a pertinent challenge, as a pilot study (Ennever, 2014a) indicates that fewer than 60% of intervocalic stop phonemes’ realizations are true stops, with the proportion dropping as low as 19% for /k/, depending on its position. By ‘reproducibility’ we refer to the challenge of reproducing another study’s results. In practice, due to the challenges of discretization and commensurability, transcription teams may invest significant resources in securing inter-coder reliability, yet in doing so, can converge upon criteria and conventions that differ from those devised in another lab. Moreover, standard instruments have their limits. Consider the stops displayed in **Figure 1**. The first appears to have a ‘break’ in F2 and F3 (cf. the analysis criteria listed in **Table 3**) while the second does not,

Table 3: Reported criteria used in stop assessments.

Source	Criteria for assessing segment as a ‘stop’
Mansfield (2015)	Significant break in vowel formants, without turbulent noise, and with some sign of a release burst in the onset of the following vowel.
Bouavichith & Davidson (2013)	A cessation of F2 and F3 during the consonant, giving rise to a period of silence (with voicing).
Marotta & Barth (2005), Ashby & Przedlacka (2011)	VOT less than half the duration of the entire segment.
Colantoni & Marinescu (2010)	Visual inspection of spectrogram.
Hualde et al. (2011)	Start marked at the end of periodic cycles of the vowel. End marked just before the burst release.
Dalcher (2006)	Total silence in the case of voiceless stops, or simply vocal fold vibration in the case of voiced stops, a visible burst, and VOT.

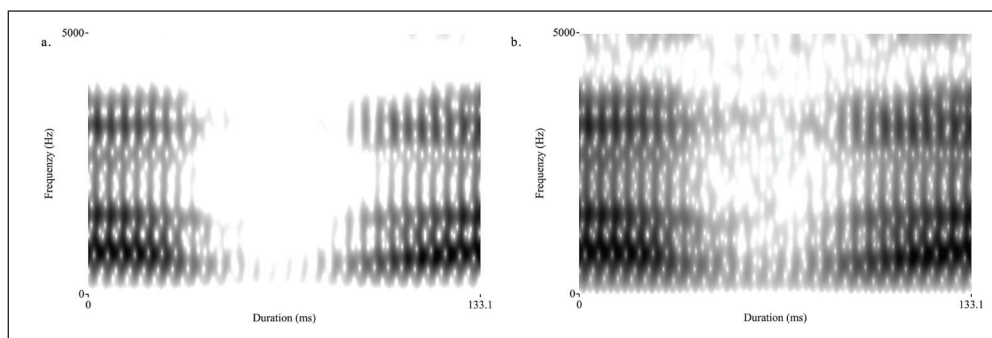


Figure 1: Stops which appear to differ in the presence of a ‘break’ in F2 and F3: **a.** is displayed with a dynamic range of 30dB and **b.** is the same stop displayed with a dynamic range of 45 dB.

yet **Figures 1a, b** depict the same token, visualized with different settings of spectrogram parameters—specifically, dynamic ranges of 30dB and 45dB respectively. Because spectrograms paint all intensities as white below some threshold, they can represent regions to the human eye as being ‘empty’ and uniform when in reality they are not, thus distorting the underlying data and inviting false comparison and analysis.

Consequently, a major contribution of this paper is methodological. In Section 3 we introduce a new method for delineating stop-like and approximant-like segments in a manner which addresses our three challenges. It uses the time-varying profile of intensity in certain frequency bands as a basis for discretizing the speech signal in terms of commensurable events (namely, threshold points in intensity velocity functions) in a fashion which is reproducible because it is automated, and deterministic given the acoustic data. Having delineated stop phonemes in this manner, we then measure the change of intensity (Δi) inside the segment, the peak intensity velocity (P_i) and the segment’s duration (D_i), each as reproducible measures of lenition and related quantities.

In previous research, measures of change of intensity (Δi) during a consonant have been employed as quantitative indexes of lenition in studies of Florentine Italian, Spanish, and American English (Bouavichith & Davidson, 2013; Colantoni & Marinescu, 2010; Dalcher, 2006; Lavoie, 2001; Lewis, 2001). Kingston (2008) and Hualde et al. (2011) in particular employ measures of peak intensity velocity (P_i) as a measure of lenition, on the grounds that more lenis variants have less abrupt acoustic transitions, making it difficult to demarcate their edges and hence determine where to measure Δi from. Thus the current study advances a line of research that infers information about lenition from careful measures of acoustic intensity. The novelty of our contribution is to couple this approach with a reproducible method for segment delineation, including of lenited variants and in a manner commensurable with the delineation of fully occluded stops, and to provide explicit arguments supporting the theoretical and empirical validity of the approach.

2. Materials

2.1. Speaker and recordings

Acoustic data are from 15 audio recordings of 1 female L1 Gurindji speaker, Violet Wadrill Nanaku. All sound files were recorded using a Roland Edirol R-09 in mono at a sample rate of 44.1 KHz with 16 bit resolution. The recordings were made by the second author between 2007–2014, when Wadrill was 66–73 years of age. The recordings consist of 14 narratives and 1 procedural-style narrative.³ The recordings were not made with acoustic analysis in mind and were recorded outside where there were some fluctuations in ambient noise levels. Generally when taking acoustic measures of intensity, it is best practice to ensure that all recording conditions are tightly controlled for, including keeping the distance between speaker and microphone constant by means of a head-mounted microphone or similar apparatus. The present study acknowledges this shortcoming but presents the following as reasons for data suitability. Firstly, the study only utilizes *relative* changes in intensity over very small time intervals (generally 0–100 ms) as measures of lenition. Absolute measures of intensity (which would be heavily impacted by any number of recording conditions) were avoided. Therefore, it was less important for the global recording conditions to remain constant and instead the central requirement was that non-vocalic sources of variation did not change significantly during the articulations under examination. Secondly, a process of audio-visual token pre-screening (detailed in Section 2.2) was employed to ensure token suitability.

³ These recordings form a part of the larger Gurindji corpus developed by the second author.

2.1.1. Gurindji's apical contrast in our recordings

Breen (2007) emphasizes that in many Australian languages, phonemic contrasts between alveolar and retroflex apicals can be elusive in the speech of some individuals. Though Gurindji has been described as possessing contrastive alveolar and retroflex apicals, the contrast in Wadrill's speech is not robust, if it exists at all. Consequently, tokens of /t/ and /t̤/ are pooled in our analysis, though we also report unpooled summary statistics in Section 5.1 (note that in word initial position, the contrast between /t/ and /t̤/ is neutralized for all speakers).

2.1.2. Qualitative features of stop lenition in Gurindji

Ennever (2014b) finds Gurindji to be typical of Pama-Nyungan languages in that it lacks fricative realizations of phonemic stops (cf. Section 1.2.1). In a qualitative analysis, he finds no evidence of frication in apical and bilabial stop articulations, and only 5 velar tokens (where $n = 208$) were found to exhibit signs of weak frication.⁴ Instead, lenition was observed to operate along a continuum that included: Fully occluded stops (**Figure 2a**), weak approximants (**Figure 2b**), more canonical approximants (**Figure 2c**) or taps (**Figure 2d**) in the case of apicals. These can be compared with a rare, weakly fricated /k/ type (**Figure 2e**).⁵

The present study focuses on quantitative measures of lenition types exemplified in the continuum as represented by **Figures (a–d)**.⁶

2.2. Initial sampling of segment tokens

Candidate phoneme tokens were identified from transcripts made by the second author, which appeared in intervocalic word-initial and intervocalic word-medial environments, i.e., V#_V and V_V. Tokens underwent audiovisual inspection in Praat (Boersma & Weenink, 2015) to ensure that they were not bounded by unexpected pauses or non-vocalic segments. Tokens that did occur in such environments, or that showed aberrant intensity profiles due to aperiodic background noise (as recordings were made outdoors) were excluded from the study. During this stage all suitable tokens were annotated at a single time point within the constriction, on a point-tier in a Praat Text-Grid. It is from this minimal markup that the automatic method described immediately below determines the boundaries of the segment and from which our relevant measures are derived.

3. An automated method for segmentation and analysis of stop phonemes

In this section we introduce an automated method for the acoustic analysis of stop phonemes, developed by the third author, which responds to the challenges of discretization, commensurability, and reproducibility identified in Section 1.4.1. We describe the method's premise (Section 3.1, Section 3.2) and the segmentation procedure (Section 3.3). We then evaluate its success and its sensitivity to parameter settings (Sections 3.4–3.6); and assess the intensity-based measures derived from the segmented data (Section 3.7). Code and documentation for the method are available online.⁷

⁴ The pre-palatal obstruent is typically realized with substantial degrees of frication in Gurindji, but we assume these reflect affricate articulatory targets, i.e., a stop + fricative sequence (cf. Section 1.1.1).

⁵ Note that the frequency viewing range is extended to 0–8000 Hz for this token.

⁶ A reviewer comments that a full-spectrum measure of intensity could conflate two distinct lenition outcomes (e.g., spirantization and stop-to-glide lenition). This is why our method focuses on specific frequency bands (see our evaluation of parameters used by the algorithm in Sections 3.4–3.7. For this study, we select bands that allow us to examine the approximant-like phones found in Gurindji, however the same algorithm with different parameter settings, and careful interpretation of their results, could be used to examine frication noise for example. This functionality allows for nuanced analysis of multiple lenition pathways.

⁷ https://github.com/erichround/stop_lenition.

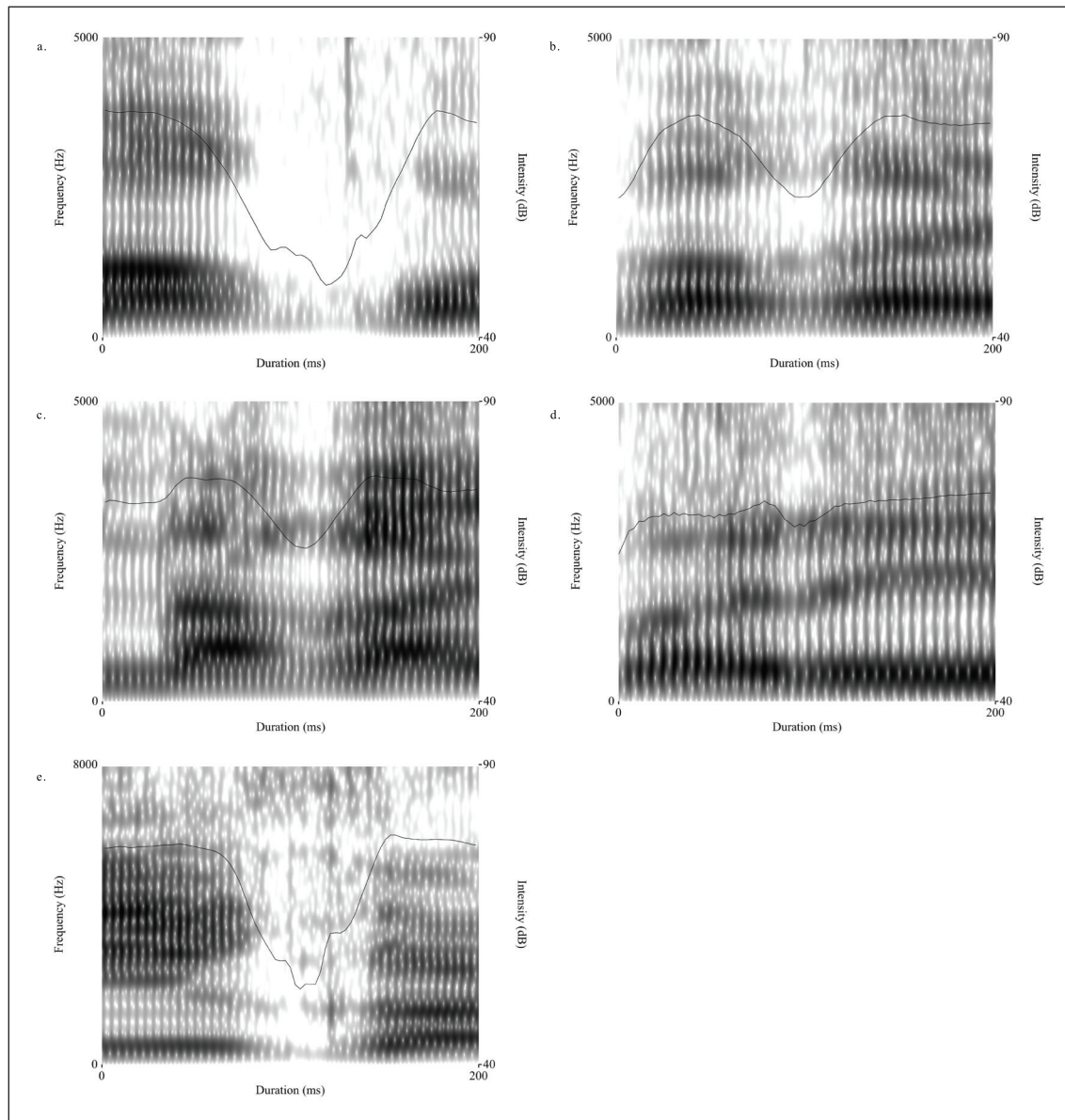


Figure 2: Spectrograms illustrating the range of stop realizations in Gurindji.

3.1. Background: Kinematic constraints on articulation

Our aim was to develop a method of interrogating acoustic data, which enables one to make meaningful inferences about articulation. Consequently, we begin with an overview of constraints on articulation. An understanding of these will help us to assess how successful the acoustic method is.

Studies of voluntary physiological movement in speech and other domains (Cooke, 1980; Munhall et al., 1985; Ostry et al., 1987) reveal tight constraints that operate on the relationships between the amplitude of a movement (A_m), its duration (D_m), and its peak velocity (P_m), which closely approximate (1), where k is constant, at least under similar speaking rates (Adams et al., 1993).

$$(1) \quad A_m = k \cdot D_m \cdot P_m$$

Equation (1) describes a three-cornered trade-off between A_m , D_m , and P_m ; for example, one might attain the same spatial magnitude of movement (A_m) while decreasing that

movement's duration (D_m) but only by increasing peak velocity (P_m); or if peak velocity is held constant, then a decrease in duration necessarily entails a decrease in movement amplitude, and so forth. True physiological systems do not match (1) exactly, but in a study of lingual and laryngeal gestures, Munhall et al. (1985) find that the basic relationship in (1) accounts for between 74% and 89% of the variance in measures of A_m , D_m , and P_m .

Our automated method makes reference to acoustic measures corresponding to A_m , D_m , and P_m . One way we would know that our method had failed to correspond well to articulation is if those acoustic measures do not closely obey an acoustic counterpart to equation (1). We apply that test in Section 3.7.

3.2. Premise of the acoustic method

The method works by delimiting segments based on acoustic data, and subsequently measuring properties of them such as duration and change of intensity.

The segments we wish to delimit are intervocalic consonants that range phonetically from true stops to more approximant-like segments (Section 2.1.2). In order to delimit these varied phonetic types in a commensurable manner, we focus on their shared articulatory properties, namely an early phase in which oral aperture decreases, and a later phase when it increases. Full closure may or may not be achieved in between. Crucially, as the aperture narrows appreciably, it causes an attenuation of the intensity of the speech signal, and thus during these focal phases, there is a broad relationship between (i) constricting/opening articulation, (ii) decreasing/increasing aperture size in the oral tract, and (iii) decreasing/increasing intensity. Consequently, to infer relative degree of constriction we measure relative intensity over time, $i(t)$. A greater total change in intensity, Δi , corresponds to narrower constriction, thus less lenition. Following practice in the processing of articulatory data, we identify landmarks for the delimitation of segments using a first derivative with respect to time, of a directly measured quantity; for our intensity function $i(t)$ we refer to that derivative as 'intensity velocity,' $v(t)$. This is described further in Section 3.3.

There are some complications we expect to encounter. In particular, some phonetic events affect intensity but are not correlated directly with oral aperture and oral constricting articulations. For segments with complete closure, passive devoicing and release bursts ought to complicate the relationship between intensity and constriction degree. Passive devoicing becomes increasingly likely as fully occluded segments become longer (Ohala, 1983)⁸ and has been described as affecting coronal stops in Tiwi (Anderson & Maddieson, 1994), an Australian language whose obstruent inventory is similar to Gurindji's. Since cessation of voicing would remove the source of sound energy, it would affect our intensity measures $i(t)$ and $v(t)$ without there being any corresponding change in the position and velocity of the superlaryngeal articulators. This may cause particularly long, fully occluded stops to have particularly large measures of Δi . Conversely, bursts at the release of a full occlusion would add a noise source that affects $i(t)$ and $v(t)$ in a manner which is separate from the effect of constriction degree. This effect may cause $i(t)$ and $v(t)$ at the right edge of a stop consonant to leap more rapidly during the

⁸ A reviewer enquired about our use of the term 'passive devoicing' in the context of stops that do not phonologically contrast voicing. We follow Anderson and Maddieson (1994) here, who use the term to describe phonetic properties of voicing in stops in Tiwi, an Australian language with an obstruent inventory similar to Gurindji, lacking a phonological voicing contrast. They note that intervocalic stop tokens—which are 'voiced' allophones—exhibited devoicing after 45–50 ms of voiced constriction. We find these same characteristics in the Gurindji data (cf. Ennever, 2014b, p. 118 ff.) and note that /k/ is more prone to devoicing than /p/ as would be expected. We tentatively conclude that—like Tiwi—there is no reason to assume that Gurindji stops are produced with an accompanying glottal opening gesture. Rather the timing and nature of a devoicing is consistent with Keating's (1984) account of passive devoicing of voiced stops in English.

burst than would be expected on the basis of superlaryngeal articulatory movement. To avoid this, our main measure of lenition will be derived from properties of the left edge of consonants.

More generally, we did not expect the relationship between intensity and articulation to hold equally well for all frequency bands in the spectrum. Higher frequencies associated with frication noise would relate to constriction in a more complex manner than we have just described. Low frequencies would also depart from the expectations described above, since they travel more readily through the walls of the vocal tract, providing in effect an acoustic side channel, whose intensity properties are not obviously linked to oral aperture and articulator position. Consequently, in designing our method we explicitly tested the utility of various frequency bands, described in Sections 3.4–3.6.

3.3. Automatic analysis and segmentation

Automatic processing was performed by custom scripts in R (R Core Team, 2016). Sound files in .WAV format were bandpassed by calling the `Filter` (pass Hann band) function of Praat (Boersma & Weenink, 2015) with a smoothing parameter of 50 Hz. Ultimately, we identified the band 400–1200 Hz to be optimal for our purposes. However, we also tested alternatives. These are assessed in Sections 3.4–3.7.

From each bandpassed sound file, a series of discrete intensity measures $\{i(t_1), i(t_2) \dots i(t_n)\}$ was extracted, with intensity analysis window of 0.01s and time step of 0.0025s, using Praat's `To Intensity` function. To this we fit a continuous, cubic spline curve $i(t)$ using `smooth.spline` (R Core Team, 2016) with the smoothing parameter $spar = 0.7$. From the continuous function $i(t)$, we calculated a first derivative with respect to time: 'intensity velocity' $v(t)$. The value 0.7 of $spar$ was chosen by experimentation, optimizing for the plausibility of the curves generated for $i(t)$ and $v(t)$; alternative values are discussed in Section 3.6.

Edges of segments were inferred from the function $v(t)$. When articulatory closure commences, intensity $i(t)$ begins to drop and intensity velocity $v(t)$ shifts rapidly to some maximum magnitude, $\max(|v(t)|)$. The demarcation algorithm uses this fact and proceeds in two steps. In our Praat TextGrid (Section 2.2) we had annotated a point somewhere within each stop, close to its beginning. The algorithm searches rightward from that 'origin' point and identifies an extremum in $v(t)$. It then delimits the left edge of the segment by selecting the moment, leading up to that extremum, when intensity velocity $v(t)$ hits a threshold level of $0.6 \cdot \max(|v(t)|)$. This demarcation point defines the beginning, not of complete closure, but of the inferred closing gesture, as intensity falls. In its second step, the algorithm searches rightwards again for the rise in $i(t)$, and associated $v(t)$ extremum, corresponding to the opening gesture. Similarly, it demarcates the start of the opening gesture using a threshold level of $0.6 \cdot \max(|v(t)|)$. Our definition of segment edges in terms of thresholds in a velocity function follows standard practice in the processing of articulatory data (cf. Kroos et al., 1997) obtained using techniques such as EMA (Schönle et al., 1987). The 60% cut-off was determined by experimentation and is evaluated in Section 3.5.

We emphasize that all segments' edges are defined in terms of the start of closing and opening gestures—properties which are shared by all of the phonetic segment types we are interested in, whether fully occluded or highly lenited. Having delimited segments in this commensurable way, we then extracted further commensurable metrics, such as its duration D_p ; the magnitude of change of intensity Δi within the segment, defined as the drop in intensity $i(t)$ from the segment's left edge to the lowest point it reaches; and peak intensity velocity P_p , defined as the greatest absolute magnitude of $v(t)$ during the segment's phase of falling intensity.

3.4. Assessing the method

Our aim was an acoustic method that is informative about articulation, and in Section 3.2 we hypothesized that some frequency bands should be more suited to this than others. In the following sections we assess various frequency bands and values of *spar*, the cubic spline smoothing parameter: We examine the algorithm's success rate for delimiting segments in Section 3.5; the quality of its delimitations in Section 3.6; the sensitivity of the derived measures D_p , Δi , and P_i to parameter choices in Section 3.7; and algebraic properties of the $i(t)$ curve in comparison to properties of articulatory movements in Section 3.8.

3.5. Success rates for segment delimitation

Our algorithm delimits segments by finding a fall–rise–fall contour in its intensity velocity profile, $v(t)$. Failures to delimit a segment can result from the absence of such a pattern in a given frequency band, or from the smoothing procedure yielding a signal which is either too noisy (insufficient smoothing) or too flat (excessive smoothing). We examined success rates of segment delimitation across nine frequency bands and four values of *spar*.

We sort our nine frequency bands into four mnemonic classes: For frequencies which predominantly carry f_0 energy, we examined two bands that we dub ‘voicing’ bands, 0–300 Hz, 0–400 Hz; for lower vocalic formants, we examined four ‘lower’ bands 300–1000 Hz, 400–1000 Hz, 400–1200 Hz, and 600–1400 Hz; for higher formants we examined two ‘upper’ bands, 1000–3200 Hz and 1200–3200 Hz; and for frication noise, one ‘noise’ band, 3200–10,000 Hz. Comparisons between band types, e.g., ‘voicing’ versus ‘lower’ should reveal which broad spectral zones provide better performance. Comparisons within band types, e.g., 300–1000 Hz versus 400–1000 Hz act as a sensitivity analysis, indicating the extent to which precise choices of upper and lower frequencies may sway our results. From the phonetic reasoning in Section 3.2 we predicted that segment delimitation using the ‘voicing’ and ‘noise’ bands would be inferior to delimitation using the ‘lower’ and ‘upper’ bands.

We compared four settings of the smoothing parameter, $spar = \{0.5, 0.6, 0.7, 0.8\}$. Given that we had already chosen $spar = 0.7$ on the basis that it produced the visually most plausible $i(t)$ and $v(t)$ functions, our prediction was that a parameter setting of 0.7 would outperform the others when we assessed it quantitatively.

Comparisons of the success rates for segment delimitation according to band choice and *spar* value are shown in **Table 4**. In this test, we ask only whether the algorithm was able to find a fall–rise–fall pattern in $v(t)$ and, on that basis, to delimit the segment. Additional questions, such as the segmentation's quality, are examined in Sections 3.6–3.8 below (in Section 3.8 we will see why $spar = 0.7$ stands out against the other *spar* values).

Success rates for segment delimitation were high in general. Comparing frequency band types, the algorithm succeeded as our phonetic reasoning predicted. Rates were

Table 4: Success rates for segment delimitation ($n = 586$ segments).

Frequency band		<i>spar</i> = 0.5	<i>spar</i> = 0.6	<i>spar</i> = 0.7	<i>spar</i> = 0.8
‘Voicing’	0–300 Hz	0.94	0.93	0.92	0.85
	0–400 Hz	0.96	0.96	0.94	0.88
‘Lower’	300–1000 Hz	0.99	0.99	0.99	0.97
	400–1000 Hz	0.99	0.99	0.99	0.97
	400–1200 Hz	0.99	0.99	0.99	0.97
	600–1400 Hz	0.98	0.99	0.99	0.97
‘Upper’	1000–3200 Hz	0.96	0.97	0.97	0.94
	1200–3200 Hz	0.96	0.98	0.97	0.90
‘Noise’	3200–10 000 Hz	0.85	0.89	0.86	0.76

highest for the ‘lower’ and ‘upper’ bands, lower for the ‘voicing’ bands, and lower again for the ‘noise’ band. Comparing within band types, the exact choice of frequency range had little effect on success rates; this suggests that the procedure is robust and is not dependent on highly specific settings of the frequency parameters. Comparing among *spar* values, the ‘lower’ bands show little variation, other than slight decline in success rates for *spar* = 0.8, due to excessive smoothing. For other band types, only *spar* = 0.8 with excessive smoothing shows any notable decline relative to the other values. This likewise indicates that the procedure is robust and is not dependent on highly specific parameter settings.

3.6. Segmentation quality

Once our algorithm finds the fall–rise–fall pattern it expects in $v(t)$, it delimits segment edges using a threshold multiple of the intensity velocity extremum. Experimentation with thresholds between $0.2 \cdot \max(|v(t)|)$ and $0.75 \cdot \max(|v(t)|)$ showed that $0.6 \cdot \max(|v(t)|)$ yielded the best results. **Figures 3a–d** display the demarcations made for a number of stop tokens with respect to their spectrograms. Smoothed intensity $i(t)$ is represented by the dotted curve, intensity velocity $v(t)$ is represented by the solid curve, and the vertical lines show the segments’ edges according to our method. Note that these demarcations do not necessarily correspond to where a human annotator would place an annotation, since whereas a human annotator will use any of a number of delimitation criteria depending

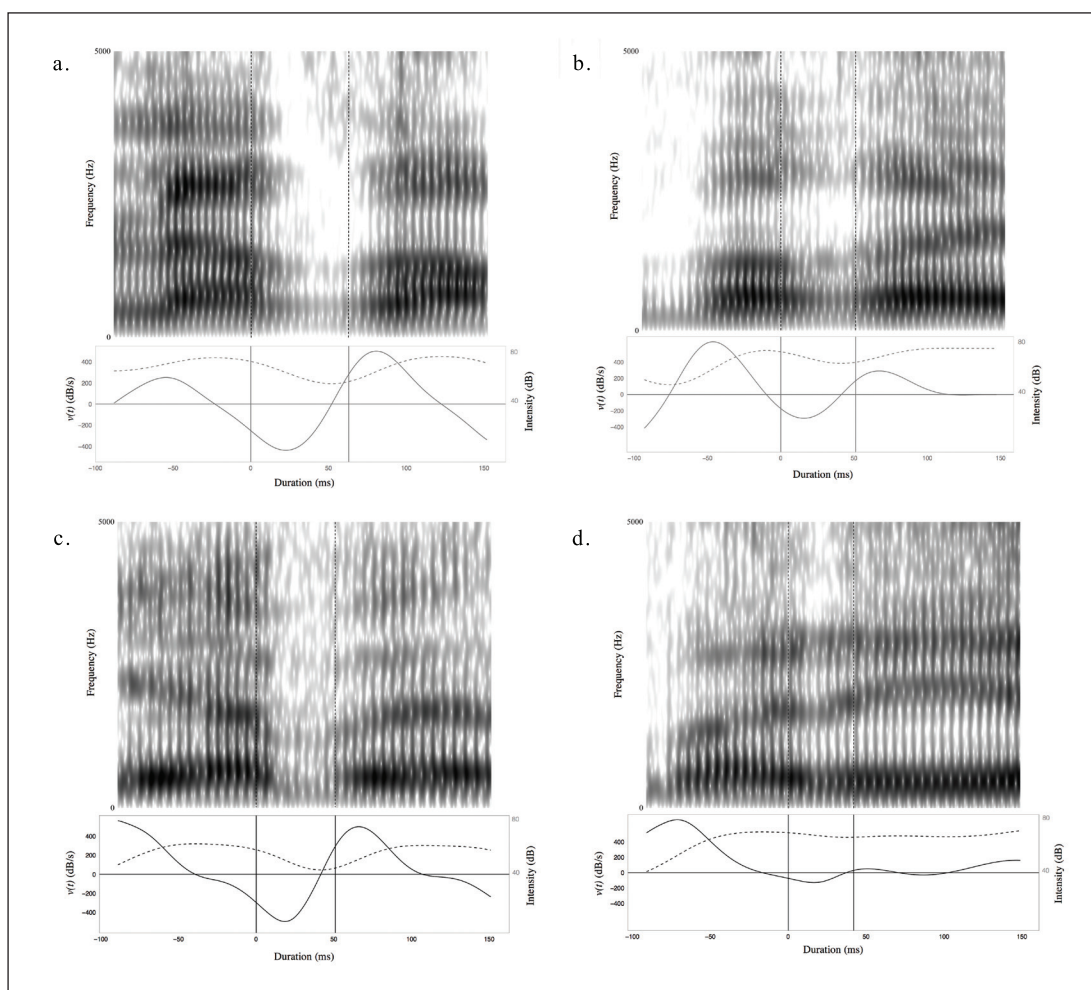


Figure 3: Example stop demarcations using *spar* = 0.7 and a delimitation threshold of $0.6 \cdot \max(|v(t)|)$ for tokens of /k/ (a, b), /p/ (c) and /t/ (d).

on the phonetic type of the token at hand, our algorithm applies the principle to all tokens, to mark beginnings of articulatory closure and opening.

Thresholds lower than $0.6 \cdot \max(|v(t)|)$ led to the left edge of segments being placed inside a preceding vowel in cases where the vowel gradually tapered in its intensity over time. Higher thresholds caused some bursts to be overlooked, leading to right edges being placed too late. Using the 400–1200 Hz frequency band and $spar = 0.7$, the algorithm using a 60% threshold delimited 581 of 586 stop phonemes. The edges it selected were manually inspected, and none were judged to be problematic.

For those segments which had bursts ($n = 112$) we also compared the position of the burst’s onset as judged by a human annotator, against the position inferred by the algorithm, using frequency band 400–1200 Hz, $spar = 0.7$, and delimitation threshold $0.6 \cdot \max(|v(t)|)$. As summarized in **Table 5**, both the mean and median differences were small, on the order of 1% of the segment’s overall duration. This indicates that for datasets of reasonable size, estimates of central tendency are of good quality. On the other hand, the standard deviation as a proportion of segment duration was 0.1, indicating that the inferred burst onset of individual tokens can differ from those judged by a human annotator. It is conceivable that the underlying cause of variation in our measurements might, for some other datasets, lead to a bias in estimates of central tendency, and we suggest that at least a subset of the inferred delimitations be compared with manual annotation, as we have done here. In future research, a customized module for better handling bursts would be a valuable addition to the method we present here.

3.7. Sensitivity of derived measures to parameter choices

In Section 3.5 we saw that exact settings of frequency bands had little effect on the algorithm’s rate of successful segment delimitation. To further evaluate our method’s sensitivity to small changes in band parameters, we compared inferred values from the ‘lower’ bands, 300–1000 Hz, 400–1000 Hz, 400–1200 Hz, and 600–1400 Hz for: Duration D_i , magnitude of change of intensity Δi , and peak intensity velocity P_i . **Table 6** presents

Table 5: Differences in burst onset position (human – automated) ($n = 112$ segments).

	Absolute difference (s)	As proportion of segment duration
Mean	0.00011	0.0015
Median	0.0011	0.014
SD	0.0081	0.10

Table 6: Measures inferred using ‘lower’ bands, compared across pairs of bands.

		Difference of means (as proportion)			Correlation, r		
		300–1000 Hz	400–1000 Hz	400–1200 Hz	300–1000 Hz	400–1000 Hz	400–1200 Hz
D_i (s)	400–1000 Hz	<i>0.01</i>			<i>0.96</i>		
	400–1200 Hz	0.02	<i>0.00</i>		0.95	<i>1.00</i>	
	600–1400 Hz	0.00	0.01	<i>0.01</i>	0.84	0.88	<i>0.89</i>
Δi (dB)	400–1000 Hz	<i>0.11</i>			<i>0.96</i>		
	400–1200 Hz	0.11	<i>0.01</i>		0.95	<i>1.00</i>	
	600–1400 Hz	0.15	0.04	<i>0.05</i>	0.83	0.87	<i>0.88</i>
P_i (dB/s)	400–1000 Hz	<i>0.14</i>			<i>0.92</i>		
	400–1200 Hz	0.13	<i>0.00</i>		0.91	<i>1.00</i>	
	600–1400 Hz	0.17	0.04	<i>0.04</i>	0.75	0.79	<i>0.80</i>

pairwise comparison of values obtained for each of the ‘lower’ bands. Comparisons shown are (i) the difference of means (expressed as a proportion of the larger of the two), which indicates the magnitude of overarching disparity, or relative bias, between bands; and (ii) linear correlation (Pearson’s r), which indicates the degree to which the disparity between a pair of bands resembles a simple, linear shift, or departs from that.

The expectation is that diagonals in **Table 6**, shown in italics, will show the lowest levels of disparity, since these compare bands that overlap the most, and this expectation is generally met. For duration D , differences of means are trivial, implying that there is little bias towards longer or shorter estimates, as the precise boundaries of the frequency bands are varied. Correlations are also high. For change of intensity, Δi , and peak intensity velocity, P , the expectation is that there will be some disparity among bands, since intensity levels in different places in the spectrum are not expected to be the same. In view of that, it is interesting that bands 400–1000 Hz and 400–1200 Hz are very similar. In sum, we find that particularly in the range of 400–1100 \pm 100 Hz, small changes to the precise band settings have little impact on derived measures of D , Δi , and P : In this part of the spectrum, our method is robust; its results are unlikely to be swayed by minor choices among possible frequency parameters.

3.8. Evaluating the method’s premise: Algebraic properties of derived measures

The premise of our method is that since articulator height correlates with oral aperture and thus with attenuation of intensity (in appropriate frequency ranges), it should be possible to use $i(t)$ as a broad proxy for articulator height and $v(t)$ for articulator velocity. If this is correct, certain algebraic properties of articulator movements (Section 3.1) should carry over to $i(t)$ and to measures based on it, D_i , and P_i . If such properties did not carry over, then this must count as evidence against the validity of our premise. Munhall et al. (1985) show duration D_m , amplitude A_m , and peak velocity P_m of articulation relate approximately as in (1). The linear relationship between A_m and $D_m.P_m$ arises when physically constrained motoric movements are optimized to minimize sudden changes in acceleration, or ‘jerk’ (Flash & Hogan, 1985; Ostry et al., 1987).

$$(1) \quad A_m = k.D_m.P_m$$

In contrast to the existence of kinematic constraints which cause articulators to obey equation (1), we are aware of no obvious equivalents, independent of articulation, which would cause acoustic measures inferred from intensity to obey equation (2), where in (2) the values D_p , Δi , and P_i are inferred from intensity.

$$(2) \quad \Delta i = k_i.D_i.P_i$$

However, if the premise of our method is sound, then we nevertheless expect equation (2) to hold, at least in those parts of the spectrum where intensity closely tracks articulation. We examine how closely our inferred measures D_p , Δi , and P_i conform to (2) in two ways. First, in Section 3.8.1 we examine correlations between Δi and $D_i.P_i$, as we vary our frequency bands and smoothing parameter *spar*. The hope is that the same parameter settings found advantageous in Sections 3.5–3.7 above are also in close accord with equation (2). Second, in Section 3.8.2 we take our best-performing parameters from Sections 3.5–3.7 and perform a full regression test to ask how closely our derived measures conform to equation (2).

3.8.1. Linear correlations

As our first test, we measure the linear correlation of Δi versus $D_i P_i$. High conformity would support our premise; low conformity would contradict it. **Table 7** shows the linear correlation (Pearson's r) of Δi and $D_i P_i$, for our nine frequency bands and four values of the cubic spline smoothing parameter $spar$. Higher correlation values indicate a closer conformity to (2), and thus by hypothesis, a closer nexus between intensity and articulation.

We interpret these results as follows. Broadly speaking, the greater the degree of smoothing applied to the underlying time series $\{i(t1), i(t2) \dots i(tn)\}$, the more the resulting continuous function $i(t)$ and its derived measures $D_p \Delta i$, and P_p , come to conform to equation (2). Interpreting this cautiously, this may arise because smoothing removes noise which otherwise obscures genuine similarities between intensity and articulation, but it may also be that smoothing reduces jerk and so coerces the data towards a function $i(t)$, whose derived measures $D_p \Delta i$, and P_p happen to have the properties in (2), or there may be an element of both. However, it can be observed that not all frequency bands are alike. Both the 'voicing' and 'noise' bands conform less well to equation (2) than the 'lower' and 'upper' band. There is no reason from the mathematics of spline fitting why this should be so, whereas the observation fits with our predictions, reasoned on phonetic grounds, regarding which bands should more closely mirror articulation. This suggests to us that spectral energy in 'lower' bands is a good choice of proxy for degree of constriction, and hence articulation.

3.8.2. Regression testing

In Section 3.8.1 we examined the relationships solely between $D_p \Delta i$, and P_i . Here we apply a more exacting test, asking also how place of articulation, neighboring vowel, position (word-internal versus -medial), and carrier word might affect that relationship. We do this by means of a linear mixed-effects regression model, with carrier word as a random effect. To be clear about what we are attempting to do here: The equation in (2) has just two degrees of freedom, so that if one specifies D_p and P_i then Δi should be fully predicted. Thus, if our acoustic measures conform to equation (2), we expect that in our regression model D_p and P_i will overwhelmingly account for the variation in Δi . If additional contributions come from the other factors, even if statistically significant, we expect their effects to be small in magnitude. If that is the case, it offers more reason to believe that our acoustic method is closely mirroring articulation.

Our regression model is summarized in **Table 8**; variables are explained below. Note that in order to keep the key terms additive, we use the equivalent of the logarithm of equation (2), $\ln(\Delta i) = k'_i + \ln(D_i) + \ln(P_i)$, where k'_i becomes an intercept term.

Table 7: Correlation of D and A/P, by frequency band and spline smoothing parameter.

Frequency band		Linear correlation, r , of Δi and $D_i P_i$			
		for $spar = 0.5$	$spar = 0.6$	$spar = 0.7$	$spar = 0.8$
'Voicing'	0–300 Hz	0.75	0.78	0.56	0.87
	0–400 Hz	0.68	0.79	0.79	0.57
'Lower'	300–1000 Hz	0.69	0.66	0.91	0.97
	400–1000 Hz	0.62	0.66	0.91	0.96
	400–1200 Hz	0.64	0.59	0.93	0.96
	600–1400 Hz	0.64	0.55	0.90	0.95
'Upper'	1000–3200 Hz	0.60	0.60	0.57	0.94
	1200–3200 Hz	0.51	0.55	0.62	0.69
'Noise'	3200–10 000 Hz	0.48	0.65	0.61	0.79

The variable Phoneme has three levels. As noted in Section 2.1.1 /t/ and /t̥/ are pooled word medially as /T/; in initial position /t̥/ does not occur. For vocalic environment, the dataset was not sufficiently large to test each preceding vowel /i,a,u/ in combination with each possible following vowel ($3 \times 3 = 9$ conditions). Instead, for each stop phoneme we binarily coded the vowel system into vowels that were/were not articulatorily proximal with each stop phoneme as per **Table 9** (see Section 1.3.3 for discussion). The resulting binary true/false values for each stop-vowel combination are provided below.

Token counts for each stop phoneme, in environments neighboring true/false proximal vowels to the left and the right, are shown in **Table 10**.

In total, 581 segments were delimited successfully (out of 586 which had been manually marked-up; see Section 2.2). Speaker was not added as a random effect because the data comes from only 1 speaker. We used a simple additive model because there were not enough data points to test interactions. The model was run using lmerTest (Kuznetsova, 2016) to test for significant predictors and MuMIn (Bartoń, 2016) to provide an R^2_c value for the model.⁹ Results are presented in **Table 11**.

As predicted, the model explains close to 100% of the variation in Δi ($R^2_c = 0.98$). It shows that the longer the duration of the stop, the greater the change of intensity Δi and hence the less likely it is to be lenited ($p < 0.001$), and similarly, the higher the peak velocity, the greater the change of intensity Δi and hence, the less likely a stop will be lenited ($p < 0.001$). The model suggests some effects of phoneme type, i.e., /p/ does not lenite to the same degree as /T/ ($p < 0.05$) and some effects for environment, i.e., a stop

Table 8: Variables potentially affecting the magnitude of Δi .

Dependant:	Log of change in intensity, $\ln(\Delta i)$	continuous (log-dB)
Fixed effects:	Log of duration, $\ln(D_i)$	continuous (log-s)
	Log of peak velocity, P	continuous (log-dB/s)
	Phoneme	categorical {/k/, /p/, /T/}
	Environment	categorical, {initial, medial}
	Proximal Preceding V	categorical, {true, false}
	Proximal Following V	categorical, {true, false}
Random effect:	Carrier Word	

Table 9: Binary variables used for vocalic environment.

	Prox = True	Prox = False
/p/	/u/	/a/, /i/
/k/	/u/	/a/, /i/
/T/	/i/	/a/, /u/

Table 10: Phoneme token counts by vocalic context.

	Preceding		Following	
	True	False	True	False
/p/	26	123	30	119
/k/	208	35	231	12
/T/	35	154	55	134

⁹ Conditional R^2 was used because it calculates variance based on both fixed and random effects and therefore takes account of all factors which are contributing to variation in the data set (Nakagawa & Schielzeth, 2013, p. 136). R^2_c was calculated using the MuMIn package in R (Bartoń, 2016).

Table 11: Summary of linear mixed effects model. REML criterion at convergence: -1456.3.

a. Scaled residuals:

Min	1Q	Median	3Q	Max
-5.9523	-0.4464	0.0689	0.4980	3.5655

b. Random effects:

Groups	Name	Variance	SD
Carrier Word	(Intercept)	0.0008455	0.02908

Number of obs: 581, groups: Carrier Word, 334.

c. Fixed effects:

	Estimate	SE	df	t value	p value
(Intercept)	-7.335972	0.070012	545.7	-104.781	<0.001
Peak Velocity	1.02745	0.008916	572.9	115.097	<0.001
Duration	0.92566	0.016038	517.6	57.676	<0.001
Phoneme /p/	-0.017768	0.008812	190.3	-2.016	<0.05
Phoneme /k/	0.011209	0.010830	220.3	1.035	0.3018
Environment	0.016245	0.007169	170.8	2.266	<0.05
Proximal preceding V	0.020114	0.007902	409.5	2.545	<0.05
Proximal following V	-0.008889	0.008688	202.2	-1.023	0.15305

is more likely to be lenited when it occurs word medially ($p < 0.05$) and after a proximal vowel ($p < 0.01$). However, as predicted, the effect sizes of each of these contributions are very small when compared to the contributions of peak velocity and duration. Although they are statistically significant, they barely contribute to accounting for Δi .

In sum, the regression analysis confirms expectations about our acoustic measures. They are behaving algebraically like the articulatory properties they are supposed to mimic. As discussed earlier, there is no inherent reason for them to do that, unless they are tracking articulation closely.

3.9. Summary and comparison with alternative ‘automated methods’

We have now introduced a quantitative method for measuring D_p , Δi , and P_i from acoustic data. The method applies commensurably to fully occluded and more lenited segments. We have tested the method and ascertained that it is highly successful at delimiting segments, delimits them in a reasonable fashion, and is not overly sensitive to small differences in parameter values. It behaves as we expected based on phonetic reasoning, and appears to mimic articulation well. Optimal settings are a frequency band of 400–1200 Hz and a *spar* parameter of 0.7.

The evidence in Section 3.7, that our acoustic measure corresponds well with articulation, accords with an explicit comparison of acoustic and articulatory measures of lenition in Spanish /b/ by Parrell (2010), which found them broadly comparable, although Parrell only investigates equivalents of our Δi vis-vis A_m ; a measure of P_i is examined but is compared not with P_m but with A_m . In combined acoustic–articulatory studies, we advocate making the comparisons we have made here.

Our method differs from existing quantitative acoustic methods, employed by Hualde et al. (2011), Carrasco et al. (2012) *inter alia*, in several respects. In Sections 3–3.7 we (i) presented an explicit phonetic rationale for why we expect our procedure to work, which relates acoustics to articulation and articulation to its kinematic constraints; (ii) assessed multiple parameters and parameter settings, and related this back to the phonetic rationale; and (iii) targeted spectral energy in frequency bands that we find most

closely mirror articulatory aperture. Our measures differ in that we focus on the changes of intensity Δi measured from the left-edge of the target consonant, where the articulation of different phonetic segment types will be largely comparable, rather than the right edge where the presence versus absence of release burst makes them less so. Our method provides a measurement of segment duration which is deterministic, and free from the variability that affects manual annotation even under the best conditions. Thus, although our focus in this paper is on intensity and lenition, we emphasize that the provision of a reproducible method for measuring duration is in itself an important methodological step forward. On a practical note, our method requires only a single manual point annotation in Praat, placed somewhere within the stop, allowing for rapid dataset mark-up requiring minimal labor and expertise.

4. Factors affecting stop lenition in Gurindji

Above, we introduced and then assessed an automated method for measuring stop lenition using acoustic data. In this section we apply the method, and enquire about the nature of gestural targets in the stop phonemes of a speaker of Gurindji. Our original research questions for Gurindji were (Section 1):

1. What is the range of realizations (in terms of lenition) of the phonemic stops in Gurindji, and their relative frequencies?
2. Are these influenced by a stop's place of articulation, preceding and proceeding vowels, and/or word boundary adjacency, and if so, how?
3. Is there evidence to support an analysis of Gurindji stop phonemes having a single, fully-occluded articulatory target, with more lenited variants the product of undershoot due to short duration; or conversely, is there evidence for more open articulatory targets also?

We first answer question (1) with standard summary statistics (Section 5.1). Our approach to answering (2) and (3) is as follows. In Section 3.8 we showed that if the aim is to predict the magnitude of change in intensity Δi , which is our measure of lenition, then it is possible to account for nearly all variation given duration D and peak intensity velocity P . So, very nearly, D and P predict Δi . But, can P itself be predicted from D ? If so, then essentially, lenition is predictable from duration alone, in accordance with the point-like model of articulatory targets (Section 1.3.5) within a Task Dynamic approach. On the other hand, if D only weakly predicts P , and therefore on its own only weakly predicts Δi , then that would accord with a more window-like interpretation of articulatory targets. Of course, there are also possible contributions from place of articulation, neighboring vowels, and boundary adjacency. Accordingly, to answer question (2) we use a second linear mixed effects model, again using the packages `lmerTest` (Kuznetsova, 2016) and `MuMIn` (Bartoń, 2016) in the R package `stats` (R Core Team, 2016). The dependent variable this time is peak intensity velocity, P , and we examine its relationships to duration, D , phoneme place of articulation, adjacency to vowels, and adjacency to boundaries. If that model can predict P with great accuracy, it will in turn predict Δi with great accuracy, and will support to point-like model (while also informing us of the relative contributions of our predictor variables). Results are presented in Section 5.2; we discuss research question (3) in Section 6.2.

4.1. Predictions

Our regression model seeks to examine the contributions to the value of P , of place of articulation, neighboring vowels, and boundary adjacency, each separate from the contribution due to duration. Any factor which increases P independently of duration would

contribute to an increase in Δi , and thus a decrease in lenition. Predictions for known factors affecting lenition are: 1. Stop phoneme tokens in word initial position will exhibit significantly less lenition (meaning we predict higher P for them) than those in word medial position (Section 1.3.4); 2. vowel quality will affect stop lenition based on co-articulation between preceding and following vowels (Section 1.3.3); and 3. /T/ may undergo less lenition (thus contribute to higher values of P) than /p/ and /k/ (Section 1.3.2).

Our predictions regarding the contribution of D to P remain more open. In an idealized scenario, if articulatory targets in Gurindji were purely point-like and there were no undershoot, then increases in duration would need to be perfectly offset by decreases in peak velocity, in order that the same target be reached consistently; this would lead to a negative relationship between D and P . In a more realistic point-like target scenario, with undershoot, we expect the relationship to be weaker than in the idealized case, but to remain negative. If targets are not point-like, then it is an open question what the operative relationship might be between D and P .

5. Results: Factors affecting stop lenition

5.1. Summary statistics

Summary statistics are in **Table 12**. **Figure 4** plots the distributions of values for D , Δi , and P by phoneme and environment. Comparison of phonemes in initial versus medial position indicate that duration is longer (higher values of D) and lenition is less pronounced (higher values of Δi) in initial positions. Duration D and degree of lenition Δi also varies across phonemes. As observed in Section 3.8.1, the linear correlation between D and Δi is significant and positive (Pearson's $r(581) = 0.68$, $p = 0.000$), i.e., the association between duration and lenition is significant and negative.¹⁰

5.2. Linear mixed effects analysis

We conducted a second linear mixed effects analysis to answer whether place of articulation, the surrounding vowels, and/or boundary adjacency have an effect on peak intensity velocity P , when one simultaneously takes into account duration D . Carrier word is added as a random effect. Speaker was not added as a random effect because the data comes from only 1 speaker. We used a simple additive model because there were not enough data points to test interactions. Results are presented in **Table 13**. The model explains a good amount of variation in the data set ($R^2_c = 0.31$).¹¹

Table 12: Summary statistics: Duration (D), change in intensity (Δi) & peak intensity velocity (P).

Position	Phoneme	N	D (s)		Δi (dB)		P (dB/s)	
			mean	SD	mean	SD	mean	SD
initial	k	172	0.068	0.015	22.19	8.44	555.4	160.8
	p	62	0.081	0.019	26.44	8.26	594.6	150.8
	T	52	0.068	0.011	19.48	4.70	512.7	120.1
medial	k	74	0.061	0.011	16.36	7.30	436.2	151.8
	p	87	0.071	0.012	22.24	6.57	557.5	135.8
	T	138	0.058	0.010	19.15	6.76	551.3	152.3
	t	84	0.059	0.010	20.17	7.09	574.8	150.0
	T	54	0.058	0.011	17.55	6.25	514.8	155.9

¹⁰ Recall that our proxy for lenition is Δi and that greater values of Δi reflect less lenition (or more stop-like articulations). Therefore, a positive correlation between duration and Δi implies a negative correlation between duration and 'degree of lenition.'

¹¹ Recall that conditional R^2 (rather than a marginal R^2) calculates variance based on both fixed and random effects and therefore takes account of all factors which are contributing to variation in the data set (Nakagawa & Schielzeth, 2013, p. 136). R^2_c was calculated using the MuMIn package in R (Bartoń, 2016).

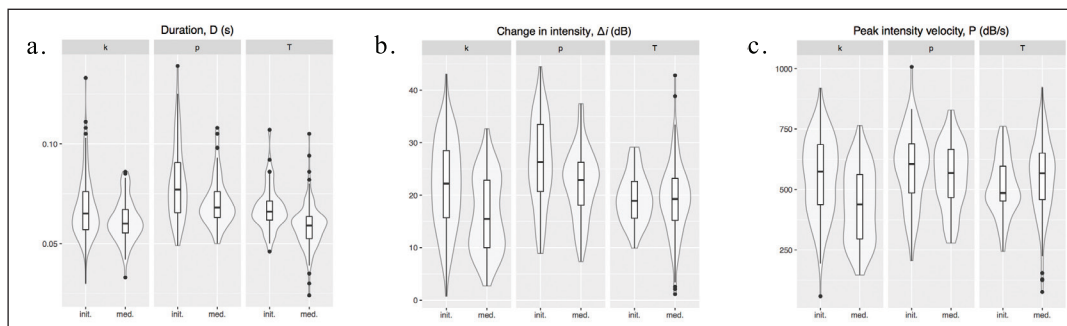


Figure 4: Distributions of duration, change in intensity and peak intensity velocity by phoneme and position. These plots combine a standard box plot, which concisely marks quantiles, with a violin plot, whose width shows the density of tokens observed across the range.

Table 13: Summary of linear mixed effects model. REML criterion at convergence: 7367.2.

a. Scaled residuals:

Min	1Q	Median	3Q	Max
-2.4533	-0.6219	0.0474	0.6679	2.4635

b. Random effects:

Groups	Name	Variance	SD
Carrier Word	(Intercept)	5047	71.04
Residual		16901	130.01

Number of obs: 581, groups: Carrier Word, 334.

c. Fixed effects:

	Estimate	SE	df	t value	p value
(Intercept)	392.41	36.28	487.40	10.816	<0.001
Duration	2818.73	472.29	551.50	5.968	<0.001
Phoneme /p/	-15.14	19.62	287.40	-0.772	0.4410
Phoneme /k /	-13.43	24.11	320.60	-0.557	0.5780
Environment	-29.55	16.05	271.10	-1.841	0.0667
Proximal preceding V	-32.68	17.46	479.50	-1.872	0.0618
Proximal following V	-1.06	19.40	299.30	-0.055	0.9565

Interestingly, the model shows a relationship between *D* and *P* which is positive: The longer the duration of the stop, the higher the peak velocity ($p < 0.001$). Neither proximal vowels, word initial versus medial environment, nor the phoneme’s place of articulation have a significant effect.

6. Discussion

6.1. Predictions versus findings

Our first prediction was that word medial position would promote greater lenition than word initial position. This is true in absolute terms; however, the absolute effect can be explained by duration (discussed below); once duration is controlled for, as in Section 5.2, we find no contribution of word-medial position to enhanced lenition compared to word-initial position. A possible confounding factor, and one not explored here, is stress (Section 1.1.2). A striking feature of many Australian languages is the cuing of stress by the lengthening of post-tonic consonants (Fletcher & Butcher, 2015). In Warlpiri (Ngumpin-Yapa), post-tonic consonants were found to be both lengthened and strengthened (Butcher & Harrington, 2003). Consequently, unlike in some other regions of the globe, it may be unusual within the Australian context for word-initial (pre-tonic) consonants to show significantly longer (and hence less lenited) stop consonants than word

medially. Importantly though, we did not distinguish post-tonic consonants from other medial consonants, nor do we have a satisfactory account of stress assignment. In any case, it would appear the results differ from Warner and Tucker's (2011) study of English in which post-stress consonants were reduced in terms of duration but *not* in terms of measures of lenition (intensity dip, cessation of formants, voicing). Further discussion of the interplay between stress, consonant lengthening, and lenition effects awaits more careful prosodic analyses of Gurindji.

Our second prediction was that flanking vowels that were articulatorily proximal to the stop constriction would be less lenited than distal flanking vowels. The model failed to confirm this. Our result instead mirrors the lack of significant effect of vocalic environment on lenition found in Murrinh Patha (Mansfield, 2015), the only other Australian language yet to be closely investigated for this effect. Taken together, these studies raise the question of how many of the widely attested lenition patterns in Australian languages are actually sensitive to vocalic environment. While preceding vowels have been demonstrated to have effects on lenition outcomes in Spanish (cf. Simonet et al., 2012; Ortega-Llebaria, 2004; Cole et al., 1999), perceptual-based models of lenition (e.g., Kingston, 2008) still contend that lenition rates are unlikely to be affected by vocalic openness. We encourage further research into the effects of vocalic environment on lenition in Australian languages to help move this debate forward.

Our third prediction was that /T/ would be less lenited than /p/ and /k/. This was not confirmed by the model. We find no evidence in Gurindji to support any of the cross-linguistic place of articulation hierarchies (cf. Section 1.3.2) claimed to govern lenition outcomes. When comparing this with our qualitative observations in Section 2.1.2 and our discussion of articulatory tapping in Section 1.3.2, we suggest that a clear division between peripheral and apical lenition is less likely to be borne out in Gurindji—and other Australian languages—once the acoustic effects of tapping are investigated quantitatively as we have done here.

6.2. Evidence for articulatory targets

Finally we return to the question of articulatory targets for phonemic stops in Gurindji. Here we emphasize that we take the results of 1 speaker only as suggestive for the possible set of phonetic facts of the Gurindji phonological system and encourage further investigation into related languages where further data acquisition is possible. Our research question was: Is there any evidence to support an analysis of Gurindji stop phonemes having a single, point-like, fully-occluded articulatory target, with more lenited variants the product of undershoot due to short duration; or conversely, is there evidence for a range of articulatory targets? If Gurindji stops have a single, fully-occluded articulatory target, as proposed for Spanish voiced stops (Parrell, 2011), then this target ought to be reached, given sufficient duration. If duration is too short, then articulation will undershoot the target, to a degree which increases as duration continues to decrease. The result, at least for undershot tokens, should be a tightly constrained, negative relationship between duration D and peak velocity P , and hence Δi . However, in Gurindji this is not the case. **Figure 5a** plots D versus Δi . The key observation is that for any horizontal cut through the data, corresponding to a single duration, one observes a wide range of Δi values, i.e., a wide variation in degree of lenition. This should only be possible if for a given duration, the speaker is making use of a range of articulatory velocities (P), which is confirmed visually in **Figure 5b**, and numerically in the R^2_c value of 0.31 for the regression model in Section 5.2 (compare $R^2_c = 0.98$ in Section 3.8.2).

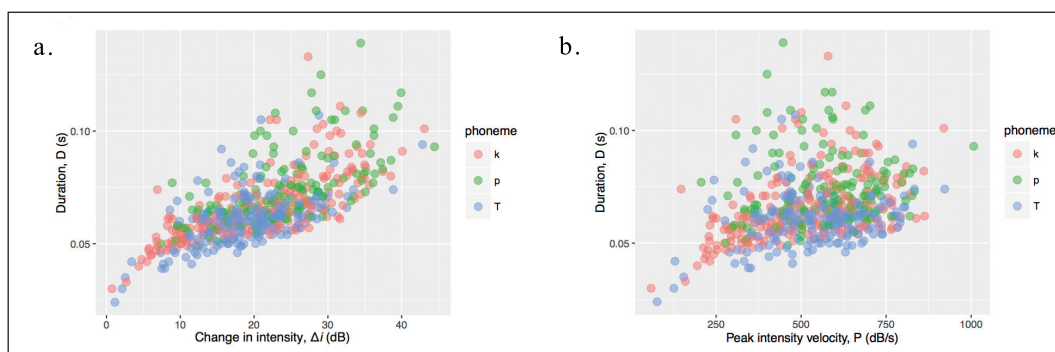


Figure 5: Change in intensity (Δi) versus **(a)** duration, **(b)** peak intensity velocity.

We can compare our results to studies that have considered the effects of ‘clock-rate’ on the realization of articulatory gestures. It is generally agreed that ‘clock slowing’¹² will cause gestures to be longer, and to have less pronounced peak velocities for a given displacement (Cho, 2001). However, if clock slowing is accompanied by larger target displacements (e.g., fully occluded stops) then these articulations will typically involve higher velocities (Byrd & Saltzman, 2003). The converse would be true if clock rate increases were accompanied by resetting of target displacements: While there is a general expectation for gestures to be shorter, or even truncated, under higher clock rates (Byrd & Saltzman, 2003), if speakers are simultaneously setting various displacement targets, then we will continue to observe a range of peak velocities. Such scenarios accord well with our observations for Gurindji. For a phonemic stop token of a given duration, the speaker in our study uses a range of velocities, some of which result in full occlusion and others which do not. We regard this as incompatible with the assumption that the speaker is aiming in all cases for a fully occluded target, since in many instances the speaker evidently could articulate more rapidly and thereby reach the target, but does not do so. One potential explanation of these observations comes from an alternative to the traditional point-attractor model of articulatory phonology, in the form of target ‘ranges’ or ‘windows.’ Such a model is formalized by Keating (1990), who proposes that gestural units are assigned individual target ‘windows’ which prescribe ranges of variability for a given articulatory dimension. We can compare our findings with Warner and Tucker (2011), who similarly argue that a ‘window’ model can account for variability in lenition phenomena in American English. They argue that conventionalized stop allophony in American English defines a reasonably broad articulatory window in which stops may be realized. Within the American English data however, the most significant constraint governing the range of these target windows—more so than mechanical/durational factors—is the preservation of phonemic distinctions, specifically between the voiced and voiceless stop series. In contrast, Gurindji, like many other Australia languages, lacks any such voicing distinction. Since there is no need to preserve a phonemic laryngeal contrast between two types of stop articulation, each stop phoneme is free to exploit a larger target window. In this context, and in the light of our findings, an interesting priority for future investigation is to improve our understanding of the extent to which articulatory targets for obstruents of Australian languages can be understood as involving a ‘window’ of targets that vary in their constriction degree, while simultaneously preserving considerable precision with respect to place of constriction.

¹² Clock-slowing temporal modulation gestures are also called prosodic gesture [π -gestures] and are proposed to have local affects in the region of phrasal boundaries.

7. Conclusions

We have measured lenition and related properties of phonemic stops in Gurindji from acoustic fieldwork data using a novel method which is precise, minimizes data-markup labor, and is automated and hence reproducible and scalable. We provided an extended evaluation of the method and found convincing evidence that our measure of lenition accords well with the properties of the articulatory system we are attempting to investigate. By attending to rates of change in intensity profiles $v(t)$, the algorithm provides deterministic, commensurate measures of segmental duration (D_i) across different phonetic realizations of phonemic stops and generates a quantitative measure of lenition (Δi).

When applied to the Gurindji data, results revealed a language whose stop phonemes span an extended space of lenition degrees, and whose patterns of lenition correspond, we argue, not to a single articulatory target but to a range, or a ‘window’ of targets encompassing both fully and partially occluded postures. Contrary to expectations, beyond the independent effect due to duration, we found no evidence of an extra positive effect on lenition due to word-medial position. The fact that word initial stops were found to be longer than their medial counterparts itself is of interest given that in Australian languages it is the post-tonic position that is typically lengthened and strengthened. Place of articulation likewise showed no significant effect—a fact that we suggest has to do with apicals freely leniting along a continuum towards taps, just as the peripheral stops lenite along a continuum towards semi-vowels.

Acknowledgements

Research was supported by Australian Research Council (ARC) grant DE150101024 and an ARC Centre of Excellence for the Dynamics of Language Small Grant to Erich Round. The collection of the Gurindji data was funded by the Jaminjung and Eastern Ngumpin DoBeS project (CI Eva Schulze-Berndt) and Endangered Languages Documentation Project (ELDP) grant IPF0134 to Felicity Meakins. We would like to thank Violet Wadrill for the use of the Gurindji recordings. We would also like to thank our two anonymous reviewers as well as Mike Proctor and audiences at the 2015 Australian Languages Society Conference for constructive feedback and commentary. Any errors and omissions remain our own.

Competing Interests

The authors have no competing interests to declare.

References

- Adams, S. G., Weismer, G. and Kent, R. D. 1993. Speaking rate and speech movement velocity profiles. *Journal of Speech, Language and Hearing Research*, 36(1), 41–54. DOI: <https://doi.org/10.1044/jshr.3601.41>
- Alpher, B. 1988. Formalizing Yir-Yoront lenition. *Aboriginal Linguistics*, 1, 188–197.
- Anderson, V. B. and Maddieson, I. 1994. Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics*, 87, 131–162.
- Ashby, M. and Przedlacka, J. 2011. The stops that aren't. *Journal of English Phonetic Society of Japan*, 1, 14–15.
- Bartoń, K. 2016. MuMin: Multi-Model Inference, CRAN.R, retrieved 03 October 2016 from: <https://CRAN.R-project.org/package=MuMin>.
- Bauer, L. 2008. Lenition revisited. *Journal of Linguistics*, 44(3), 605–624. DOI: <https://doi.org/10.1017/S0022226708005331>
- Blevins, J. 2001. Where have all the onsets gone? Initial consonant loss in Australian Aboriginal languages. In: Simpson, J., Nash, D., Laughren, M., Austen, P. and Alpher, B. (eds.), *Forty years on: Ken Hale and Australian languages*, 481–492.

- Boersma, P. and Weenik, D. 2015. Praat: Doing phonetics by computer [Computer Program]. Version 6.0.20, retrieved 3 September 2015 from: <http://www.praat.org/>.
- Bouavichith, D. and Davidson, L. 2013. Segmental and prosodic effects on intervocalic voiced stop reduction in connected speech. *Phonetica*, 70(3), 182–206. DOI: <https://doi.org/10.1159/000355635>
- Breen, G. 1997. Taps, stops and trills. In: Tyron, D. and Walsh, M. (eds.), *Boundary Rider: Essays in honour of Geoffery O'Grady*, 71–93. Canberra: Pacific Linguistics.
- Breen, G. 2007. Reflecting on Reflexion. *Paper presented at OzPhon07: Workshop on the phonetics and phonology of Australian languages*, La Trobe University, 3–4 December 2007.
- Browman, C. P. and Goldstein, L. 1989. Articulatory gestures as phonological units. *Phonology*, 6(2), 201–251. DOI: <https://doi.org/10.1017/S0952675700001019>
- Bundgaard-Nielsen, R. L., Baker, B. J., Kroos, C., Harvey, M. and Best, C. T. 2012. Vowel acoustics reliably differentiate three coronal stops of Wubuy across prosodic contexts. *Lab Phon*, 3, 133–161. DOI: <https://doi.org/10.1515/lp-2012-0009>
- Bundgaard-Nielsen, R. L., Baker, B. J., Kroos, C., Harvey, M. and Best, C. T. 2015. Discrimination of multiple coronal stop contrasts in Wubuy (Australia): A natural referent consonant account. *PLoS One*, 10(12), e0142054. DOI: <https://doi.org/10.1371/journal.pone.0142054>
- Busby, P. A. 1980. The distribution of phonemes in Australian Aboriginal Languages. *Papers in Australian Linguistics*, 4, 73–139.
- Butcher, A. 1995. The phonetics of neutralisation: The case of Australian coronals. In: Windsor-Lewis, J. (ed.), *Studies in general and English phonetics: Essays in honour of Professor JD O'Connor*, 10–38. New York: Routledge.
- Butcher, A. 2004. Fortis/lenis revisited one more time: The aerodynamics of some oral stop contrasts in three continents. *Clinical Linguistics & Phonetics*, 18(6), 547–557. DOI: <https://doi.org/10.1080/02699200410001703565>
- Butcher, A. 2006. Australian aboriginal languages: Consonant-salient phonologies and the 'place-of-articulation imperative.' In: Harrington, J. and Tabain, M. (eds.), *Speech production: Models, phonetic processes, and techniques*, 187–210. New York: Psychology Press.
- Butcher, A. and Harrington, J. 2003. An acoustic and articulatory analysis of focus and the word/morpheme boundary distinction in Warlpiri. In: Palethorpe, S. and Tabain, M. (eds.), *Proceedings of the 6th International Seminar in Speech Production*, 19–24. Sydney: Macquarie University.
- Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language variation and change*, 14(3), 261–290. DOI: <https://doi.org/10.1017/S0954394502143018>
- Byrd, D. and Saltzman, E. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149–180. DOI: [https://doi.org/10.1016/S0095-4470\(02\)00085-2](https://doi.org/10.1016/S0095-4470(02)00085-2)
- Carrasco, P. G., Hualde, J. I. and Simonet, M. 2012. Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica*, 69(3), 149–179. DOI: <https://doi.org/10.1159/000345199>
- Chadwick, N. J. 1975. *A descriptive study of the Djingili language*, 2. Canberra: Australian Institute of Aboriginal Studies.
- Cho, T. 2001. Effects of prosody on articulation in English (Doctoral dissertation). Los Angeles, CA: University of California, Los Angeles.
- Cho, T. 2006. Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In: Goldstein, L., Whalen, D. H. and Best, C. (eds.), *Laboratory phonology 8: Varieties of phonological competence*, 519–548. Berlin: Mouton de Gruyter.

- Colantoni, L. and Marinescu, I. 2010. The scope of stop weakening in Argentine Spanish. In: Ortega-Llebaria, M. (ed.), *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, 100–114. Somerville, MA: Cascadilla Proceedings Project.
- Cole, J., Hualde, J. I. and Iskarous, K. 1999. Effects of prosodic and segmental context on /g/-lenition in Spanish. In: Fujimura, O., Joseph, B. D. and Palek, B. (eds.), *Proceedings of the 4th international linguistics and phonetics conference, 2*, 575–589. Prague: The Karolinium Press.
- Connell, B. 1991. Phonetic aspects of the Lower Cross languages and their implications for sound change (Doctoral dissertation). Edinburgh: University of Edinburgh.
- Cooke, J. D. 1980. The organization of simple, skilled movements. In: Stelmach, G. and Requin, J. (eds.), *Tutorials in motor behavior*, 199–212. Amsterdam: North-Holland. DOI: [https://doi.org/10.1016/S0166-4115\(08\)61946-9](https://doi.org/10.1016/S0166-4115(08)61946-9)
- Dalcher, C. 2006. Consonant weakening in Florentine Italian: An acoustic study of gradient and variable sound change (Doctoral dissertation). Washington, D.C.: Georgetown University.
- Dixon, R. M. W. 2002. Phonology. In: Dixon, R. M. W. (ed.), *Australian languages: their nature and development, 1*, 547–658. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486869.015>
- Donegan, P. J. and Stampe, D. 1979. The study of natural phonology. In: Dinnsen, A. D. (ed.), *Current approaches to phonological theory*, 126–173. Bloomington: Indiana University Press.
- Edwards, J., Beckman, M. E. and Fletcher, J. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89(1), 369–382. DOI: <https://doi.org/10.1121/1.400674>
- Ennever, T. 2014a. A closer look at the phonetic variation of Gurindji obstruents. *Paper presented at Australian Linguistics Society Annual Meeting*, Newcastle Australia, 11 December 2014.
- Ennever, T. 2014b. Stop lenition in Gurindji: An acoustic phonetic study (Honours dissertation). St. Lucia: The University of Queensland, St. Lucia.
- Escure, G. 1977. Hierarchies and phonological weakening. *Lingua*, 43, 55–64. DOI: [https://doi.org/10.1016/0024-3841\(77\)90048-1](https://doi.org/10.1016/0024-3841(77)90048-1)
- Evans, B. and Merlan, F. 2004. Stop contrasts in languages of Arnhem Land: From the perspective of Jawoyn, Southern Arnhem Land. *Australian Journal of Linguistics*, 24(2), 185–224. DOI: <https://doi.org/10.1080/0726860042000271825>
- Evans, N. 1995. Current issues in the phonology of Australian languages. In: Goldsmith, J. A. (ed.), *The handbook of phonological theory*, 723–761. Massachusetts: Blackwell Publishing.
- Fant, G. 1973. *Speech sounds and features*. Cambridge, MA: MIT Press.
- Flash, T. and Hogan, N. 1985. The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, 5(7), 1688–1703.
- Fletcher, J. and Butcher, A. 2015. Sound patterns of Australian languages. In: Koch, H. and Nordlinger, R. (eds.), *The languages and linguistics of Australia, 3*, 91–138. Berlin: Walter de Gruyter.
- Foley, J. 1977. *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.
- Fry, D. B. 1979. *The physics of speech*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139165747>

- Guenther, F. H. 1995. Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. *Psychological Review*, 102, 594–621. DOI: <https://doi.org/10.1037/0033-295X.102.3.594>
- Gurevich, N. 2008. *Lenition and contrast*. London: Routledge.
- Hualde, J. I. and Nadeu, M. 2011. Lenition and phonemic overlap in Rome Italian. *Phonetica*, 68(4), 215–242. DOI: <https://doi.org/10.1159/000334303>
- Hualde, J. I., Simonet, M. and Nadeu, M. 2011. Consonant lenition and phonological recategorisation. *Laboratory phonology*, 2(2), 301–329. DOI: <https://doi.org/10.1515/labphon.2011.011>
- Ingram, J., Laughren, M. and Chapman, J. 2008. Connected speech processes in Warlpiri. *Paper presented at the Interspeech 2008: 9th Annual Conference of the International Speech Communication Association: 2008*, 1–5.
- Kaplan, A. 2010. Phonology shaped by phonetics: The case of intervocalic lenition (Doctoral dissertation). Santa Cruz, CA: University of California, Santa Cruz.
- Keating, P. 1984. Phonetic and phonological representation of stop consonant voicing. *Language*, 286–319. DOI: <https://doi.org/10.2307/413642>
- Keating, P. 1990. The window model of coarticulation: Articulatory evidence. In: Kingston, J. and Beckman, M. (eds.), *Papers in Laboratory Phonology I*, 451–470. Cambridge: Cambridge University Press.
- Kingston, J. 2008. Lenition. In: Colantoni, L. and Steele, J. (eds.), *Selected proceedings of the 3rd conference on laboratory approaches to Spanish phonology*, 1–31. Somerville, MA: Cascadilla Proceedings Project.
- Kirchner, R. 2001. An effort based approach to consonant lenition (Doctoral dissertation). London: Routledge.
- Kirchner, R. 2004. Consonant lenition. In: Hayes, B., Kirchner, R. and Steriade, D. (eds.), *Phonetically based phonology*, 313–345. Oxford: Oxford University Press. DOI: <https://doi.org/10.1017/CBO9780511486401.010>
- Koch, H. 2004. The Arandic subgroup of Australian languages. In: Bower, C. and Koch, H. (eds.), *Australian languages: Classification and the comparative method*, 127–150. Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/cilt.249.10koc>
- Kroos, C., Hoole, P., Kühnert, B. and Tillmann, H. G. 1997. Phonetic evidence for the phonological status of the tense-lax distinction in German. *Forschungsberichte- Institut für Phonetik und Sprachliche Kommunikation der Universität München*, 35, 17–25.
- Kuznetsova, A. 2016. *lmerTest: Tests in Linear Mixed Effect Models: CRAN.R*, retrieved 03 October 2016 from: <https://CRAN.R-project.org/package=lmerTest>.
- Ladefoged, P. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Wiley-Blackwell.
- Lass, R. 1984. *Phonology*. Cambridge: Cambridge University Press.
- Lavoie, L. M. 2001. *Consonant strength: Phonological patterns and phonetic manifestations*. New York: Routledge.
- Lewis, A. 2001. Weakening of Intervocalic /p, t, k/ in Two Spanish Dialects: Towards the Quantification of Lenition Processes (Doctoral dissertation). Urbana-Champaign, IL: University of Illinois, Urbana-Champaign.
- Lindblom, B. 1983. *Economy of speech gestures*. New York: Springer. DOI: https://doi.org/10.1007/978-1-4613-8202-7_10
- Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W. J. and Marchal, A. (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-009-2037-8_16

- Mansfield, J. B. 2015. Consonant lenition as a sociophonetic variable in Murrinh Patha (Australia). *Language Variation and Change*, 27(02), 203–225. DOI: <https://doi.org/10.1017/S0954394515000046>
- Marotta, G. and Barth, M. 2005. Acoustic and sociolinguistic aspects of lenition in Liverpool English. *Studi Linguistici e Filologici Online*, 3(2), 377–341.
- Marslen-Wilson, W. and Zwisterlood, P. 1989. Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585. DOI: <https://doi.org/10.1037/0096-1523.15.3.576>
- McConvell, P. and Meakins, F. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics*, 25(1), 9–30. DOI: <https://doi.org/10.1080/07268600500110456>
- McGregor, W. 1990. *A functional grammar of Gooniyandi*. Amsterdam: John Benjamins Publishing. DOI: <https://doi.org/10.1075/slcs.22>
- McKay, G. R. 1980. Medial stop gemination in Rembarrnga: A spectrographic study. *Journal of Phonetics*, 8, 343–352.
- Meakins, F., McConvell, P., Charola, E., McNair, N., McNair, H. and Campbell, L. 2013. *Gurindji to English Dictionary*. Batchelor, Australia: Batchelor Press.
- Munhall, K. G., Ostry, D. J. and Parush, A. 1985. Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance*, 11(4), 457–474. DOI: <https://doi.org/10.1037/0096-1523.11.4.457>
- Nakagawa, S. and Schielzeth, H. 2013. A general and simple method for obtaining R2 from general linear mixed-effects models. *Methods in Ecology and Evolution* 4(2), 133–142. DOI: <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. *The production of speech*, 189–216. New York: Springer-Verlag. DOI: https://doi.org/10.1007/978-1-4613-8202-7_9
- Ohala, J. J. 1996. The relation between sound change and connected speech processes. *Arbeitsberichte*, 31, 201–206.
- Oller, K. D. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54(5), 1235–1247. DOI: <https://doi.org/10.1121/1.1914393>
- Ortega-Lleberia, M. 2004. Interplay between phonetic and inventory constraints in the degree of spirantization of voiced stops: Comparing intervocalic /b/ and intervocalic /g/ in Spanish and English. In: Face, T. L. (ed.), *Laboratory approaches to Spanish phonology*, 237–253. Berlin: Mouton de Gruyter.
- Ostry, D. J., Cooke, J. D. and Munhall, K. G. 1987. Velocity curves of human arm and speech movements. *Experimental Brain Research*, 68(1), 37–46. DOI: <https://doi.org/10.1007/BF00255232>
- Parrell, B. 2010. Articulation from acoustics: Estimating constriction degree from the acoustic signal. *Journal of the Acoustical Society of America*, 128(4), 2289. DOI: <https://doi.org/10.1121/1.3508033>
- Parrell, B. 2011. Dynamical account of how /b, d, g/ differ from /p, t, k/ in Spanish: Evidence from labials. *Laboratory phonology*, 2(2), 423–449. DOI: <https://doi.org/10.1515/labphon.2011.016>
- Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J. and Hopper, P. *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.45.08pie>

- Proctor, M., Bundgaard-Nielsen, R. L., Best, C., Goldstein, L., Kroos, C. and Harvey, M. 2010. Articulatory modelling of coronal stop contrasts in Wubuy. *Proceedings of the 13th Australian international conference of speech science and technology*, 90–93.
- R Core Team. 2016. R: A language and environment for statistical computing. Vienna: R Foundation for statistical computing. Retrieved from: <http://www.R-project.org>.
- Round, E. R. 2010. Widespread patterns of lenition in Australian indigenous languages. *13th Australasian International Conference on Speech Science and Technology, Melbourne*.
- Saltzman, E. L. and Munhall, K. G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological psychology*, 1(4), 333–382. DOI: https://doi.org/10.1207/s15326969eco0104_2
- Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J. and Conrad, B. 1987. Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31(1), 26–35. DOI: [https://doi.org/10.1016/0093-934X\(87\)90058-7](https://doi.org/10.1016/0093-934X(87)90058-7)
- Ségéral, P. and Scheer, T. 2008. Positional factors in lenition and fortition. In: de Carvalho, C., Scheer, T. and Ségéral, P. (eds.), *Lenition and fortition*, 131–172. Berlin: Mouton de Gruyter.
- Shockey, L. and Gibbon, F. 1993. “Stopless stops” in connected English. *Speech Research Laboratory, University of Reading Work in Progress*, 7, 1–7.
- Simonet, M., Hualde, J. I. and Nadeu, M. 2012. Lenition of /d/ in spontaneous Spanish and Catalan. *Interspeech Proceedings*, 1416–1419.
- Soler, A. and Romero, J. 1999. The role of duration in stop lenition in Spanish. In: *Proceedings of the 14th International Congress of Phonetic Sciences*, 483–486.
- Steriade, D. 2001. Directional asymmetries in place assimilation. In: Hume, B. and Johnson, K. (eds.), *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.
- Stoakes, H., Fletcher, J. and Butcher, A. 2007. An acoustic study of Bininj Gun-Wok medial stop consonants. *Paper presented at the Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken: University of Saarland.
- Tabain, M. 2003. Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America*, 113, 2834–2849. DOI: <https://doi.org/10.1121/1.1564013>
- Tabain, M. 2009. An EPG study of the alveolar vs. retroflex apical contrast in Central Arrernte. *Journal of Phonetics*, 37(4), 486–501. DOI: <https://doi.org/10.1016/j.wocn.2009.08.002>
- Tabain, M. and Butcher, A. 2015. Stop bursts in Pitjantjatjara. *Journal of the International Phonetic Association*, 45(02), 149–176. DOI: <https://doi.org/10.1017/S0025100315000110>
- Tabain, M. and Rickard, K. 2007. A preliminary EPG study of stop consonants in Arrernte. *Paper presented at the Proceedings of the 16th International Congress of Phonetic Sciences*.
- Turk, A., Nakai, S. and Sugahara, M. 2006. Acoustic segment durations in prosodic research: A practical guide. In: Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. and Schließer, J. (eds.), *Methods in empirical prosody research*, 3, 1–28. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110914641.1>
- Warner, N. and Tucker, B. V. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3), 1606–1617. DOI: <https://doi.org/10.1121/1.3621306>
- Wheeler, M. 2005. *The phonology of Catalan*. Oxford: Oxford University Press.

- Wood, R. 1978. Some Yuulngu phonological patterns. In: Kirton, J. F. and Australian National University (eds.), *Papers in Australian Linguistics*, 11, 53–118. Canberra: Pacific Linguistics.
- Zwicky, A. 1972. On casual speech. *Paper presented at the Eighth Regional Meeting of the Chicago Linguistic Society*.

How to cite this article: Ennever, T., Meakins, F. and Round, E. R. 2017 A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 20, pp.1–32, DOI: <https://doi.org/10.5334/labphon.18>

Submitted: 15 April 2016 **Accepted:** 01 May 2017 **Published:** 22 August 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 