

JOURNAL ARTICLE

Inter-consonantal intervals in Tripolitanian Libyan Arabic: Accounting for variable epenthesis

Leendert Plug, Abdurraouf Shitaw and Barry Heselwood

School of Languages, Cultures and Societies, University of Leeds, UK

Corresponding author: Leendert Plug (L.plug@leeds.ac.uk)

This paper reports on an acoustic investigation of inter-consonantal intervals in plosive sequences in Tripolitanian Libyan Arabic (TLA). TLA permits a wide range of two, three, and four-consonant strings within and across word boundaries. Previous descriptive work has suggested that TLA is characterized by widespread, partly optional vowel epenthesis throughout these sequences. We conducted a production study to investigate the phonetic and phonological properties of inter-consonantal intervals (ICIs) in these sequences, including their durations and voicing characteristics, as well as the voicing of surrounding plosive hold phases. Our aim was to assess the extent of ‘variable epenthesis’ and narrow down the range of possible accounts for it. Unlike many previous studies, we did not rely on native speaker intuitions and made minimal use of manual categorical coding, instead focusing on modelling distributions of continuous acoustic parameters. The central question we address is whether our data contain patterns that can only be explained with reference to phonological vowel insertion—given that variable epenthesis can also be accounted for in terms of vocoid intrusion alone (Hall, 2006). Our study improves our understanding of consonant articulations in Arabic, and of epenthesis and related phenomena.

Keywords: epenthesis; vowel intrusion; voicing assimilation; Libyan Arabic

1. Introduction

This paper reports on an acoustic investigation of inter-consonantal intervals (henceforth ‘ICIs’) in plosive sequences in the Arabic dialect spoken in the capital of Libya, Tripoli. Tripolitanian Libyan Arabic (henceforth ‘TLA’) is of interest because it permits a wide range of two, three, and four-consonant strings within and across word boundaries, and previous descriptive work has suggested TLA is characterized by widespread, partly optional vowel epenthesis throughout these sequences. Our study was motivated by an interest in the phonetic details of, and phonological constraints on, the resulting variation.

Like most other Arabic dialects, TLA does not permit clusters of more than two consonants within words; therefore, three-consonant and four-consonant sequences across word boundaries are phonologically marked. The typical result of this markedness, at least in Arabic dialects, is the appearance of an epenthetic vowel in these sequences, breaking a three-consonant string up into a single C plus a CC cluster or sequence, and breaking a four-consonant sequence up into two CC clusters. This phonologically motivated process of epenthesis is well-documented (Owens, 2006; Watson, 2002), and has given rise to a typology of Arabic dialects (Broselow, 1992; Kiparsky, 2003; Watson, 2007). Watson (2007) describes TLA as a ‘VC’ dialect in this typology. This reflects the observation that in sequences of three consonants, epenthesis occurs between the first two consonants, irrespective of the position of a morpheme or word boundary: -CvC#C- and -Cv#CC-, where ‘v’ signifies an epenthetic vowel and ‘#’ a morpheme or word boundary. In other

Arabic dialects an epenthetic vowel systematically occurs after the middle consonant ('CV' dialects), or no epenthesis occurs ('C' dialects). In four-consonant sequences, 'VC' dialects have epenthesis at the morpheme or word boundary: -CCv#CC-.

Following Watson's description of TLA, this phonologically motivated epenthesis process yields three 'sites' in which epenthesis is expected to be widespread if not obligatory: -CvC#C-, -Cv#CC- and -CCv#CC-. However, Watson (2007, p. 345) adds the observation that in fact, 'all permissible clusters may be broken up by epenthesis': among the examples she provides is /xubz/ [xubz]~[xubəz] 'bread.' Watson reports that "epenthesised and non-epenthesised forms are not social or geographical variants, although they may well be stylistic variants; both may be used by one and the same speaker." In other words, TLA vowel epenthesis is more variable than a simple account in terms of phonotactics might suggest. A detailed analysis of its phonetics is therefore warranted.

1.1. Accounting for variable epenthesis

Variability in surface forms that may be derived from a phonological process can be accounted for in a number of ways, depending on the nature of the variability (Bayles, Kaplan, & Kaplan, 2016; Bellik, 2018; McPherson & Hayes, 2016). In the case of variable vowel epenthesis patterns, a crucial issue concerns the relative contributions of phonologically motivated vowel insertion and phonetically motivated variability in inter-consonantal timing. Taken at face value, Watson's (2007) account of variable epenthesis in TLA suggests that epenthesis is a categorical process which applies obligatorily, or near-obligatorily, in three inter-consonantal sites, and optionally—with intra-speaker variation and possible stylistic constraints—in other inter-consonantal sites. This means that the sites are distinguished, within and possibly across speakers, by the associated frequencies of occurrence of epenthetic vowels.

Watson's account is similar to that of variable epenthesis in Moroccan Arabic as proposed by Heath (1987). Heath's account has been challenged by Dell and Elmedlaoui (2002) and Gafos (2002), who suggest that the 'epenthetic vowels' of Moroccan Arabic are better characterized as 'intrusive vocoids'—also called 'excrecent,' 'emergent,' and 'transitional' vocoids (Gafos, 2002; Hall, 2006, 2011). Intrusive vocoids result from inter-consonantal coordination patterns characterized by a low degree of overlap between the gestures associated with the two consonants, or 'open transition' (Browman & Goldstein, 1992; Catford, 2001; Gafos, Hoole, Roon, & Zeroual, 2010). While short open transitions are hearable as first-consonant releases, longer open transitions can be perceived as vowel-like segments between the two consonants, especially if they are voiced as a result of variability in the coordination of vocal fold vibration associated with one or both of the consonants (Gafos, 2002; Shaw & Davidson, 2011). In this account, voiced inter-consonantal intervals do not arise through a phonological process of vowel insertion; different phonological contexts are distinguished by their associated ranges of variability in inter-consonantal timing.

The question of whether given data patterns are best accounted for in terms of a phonological process of vowel insertion or in terms of variable inter-consonantal coordination has been asked in the study of multiple languages. For example, Ridouane (2008) and Ridouane and Fougeron (2011) argue that epenthesis in Tashlhiyt Berber is best accounted for in terms of vocoid intrusion, contra Coleman (2001). Bellik (2018) argues the same for apparent epenthesis in Turkish onset clusters, contra Clements and Sezer (1982). However, some consideration suggests that the choice between these two alternative accounts, which we might call the 'epenthesis account' and the 'intrusion account', respectively, is not necessarily a simple one to make in languages in which non-final consonants in CC, CCC, and CCCC sequences are typically realized with audible

release—that is, languages whose consonant sequences are generally characterized by open transition. Moroccan Arabic is among these languages (Gafos et al., 2010), and so is TLA (Ghummed, 2015; Shitaw, 2014). Obligatory or near-obligatorily open transition means that there will be some degree of variation in the duration of the transition, providing a temporal window in which intrusive vocoids can emerge. This in turn means that an epenthesis account ruling out any vocoid intrusion is unlikely to be descriptively adequate: A more realistic version of this account is one in which the language is characterized as having both epenthetic vowels and intrusive vocoids, as has been suggested for Lebanese Arabic (Hall, 2013). In contrast, an intrusion account ruling out any phonological vowel insertion might in principle provide a good fit for observed data patterns.

For languages like TLA, a first question to ask is therefore whether its variable epenthesis comprises data patterns that cannot be straightforwardly accounted for in terms of interacting factors influencing inter-consonantal timing, without the postulation of a phonological vowel insertion process. If the postulation of a vowel insertion process seems necessary, a next question is what the relative contributions of vocoid intrusion and vowel epenthesis are to the observed data patterns. Our aim in this investigation was to assess whether detailed consideration of the temporal and voicing characteristics of inter-consonantal intervals might allow us to narrow down the range of potential accounts for variable epenthesis in TLA.

1.2. Evidence for phonological vowel insertion

In order to address the question of whether TLA variable epenthesis requires reference to phonological vowel insertion, we need to be clear on what counts as evidence for a phonological epenthesis rule. A review of the literature, including Hall's (2006) cross-linguistic survey of vowel epenthesis phenomena, suggests a number of data patterns can be used in support of an account involving vowel insertion.

First, Hall's (2006) survey shows that vowels accounted for in terms of phonological epenthesis tend to be consistently voiced, whereas intrusive vocoids alternate with voiceless inter-consonantal intervals. Epenthetic vowels also tend to be described as similar to lexical vowels in showing clear formant structure in higher frequency ranges, while voiced open transitions can show low-level voicing similar to that of 'voicing tails' in plosive hold phases. This makes sense in that inserted vowels can be expected to be realized similarly to lexical vowels, and should not be *a priori* prone to devoicing. Given that intrusive vocoids emerge out of variation in the temporal coordination of articulatory gestures that include, but are not restricted to laryngeal gestures, their voicing is both context-dependent—in that its source is in the surrounding segments—and variable—in that voicing in surrounding segments is not necessarily carried through the inter-consonantal interval. Notably, an observation of consistently voiced inter-consonantal intervals with formant structure between voiceless consonants is difficult to account for in terms of vocoid intrusion, and therefore constitutes evidence in favour of an account involving phonological vowel insertion.

Second, according to Hall (2006), a review of studies presenting duration measurements for epenthetic vowels and intrusive vocoids warrants the generalization that the former are longer than the latter. This makes sense in that all other things being equal, vowels with a dedicated tongue-body gesture can be expected to take longer to produce than vowel-like segments resulting from variation in the timing of adjacent consonantal gestures. Several studies appeal to a related diagnostic concerning the duration of the consonant cluster in which epenthesis is observed: While epenthetic vowels add significantly to this duration, intrusive vocoids do not (Ridouane & Fougeron, 2011). In a language like TLA,

in which variable open transition is expected, bimodal distributions of inter-consonantal interval and cluster durations can be taken as evidence in favour of an account involving phonological vowel insertion—unless the bimodality can be attributed to the interplay of independently motivated constraints on inter-consonantal timing.

Third, Hall (2006) points out that intrusive vocoids are generally transparent to phonological processes, while epenthetic vowels tend to be phonologically ‘visible.’ This is reflected in the fact that epenthesis processes feature prominently in the phonological literature on rule interactions. For example, Davis (1995) observes that in southern Palestinian Arabic, rightward assimilation of emphasis is blocked by epenthetic vowels: /bat^hnak/ ‘your (m.sg.) stomach’ is realized [b^ha^ht^hn^ha^hk^h], with both leftward and rightward emphasis spread, while /bat^hnha/ ‘her stomach’ is realized [b^ha^ht^hinha], with an epenthetic vowel [i] and leftward emphasis spread only. In a rule-based account, epenthesis operates before assimilation, removing the context for the latter to apply in. Odden (2013) appeals to the opposite order to explain the transparency of Armenian epenthetic vowels to voicing assimilation, but these ‘epenthetic vowels’ are better characterized as intrusive vocoids (Hall, 2006; Vaux, 2003). In TLA, in which (near-) obligatory epenthesis is expected in some inter-consonantal sites but not others, an observation of significantly lower rates of process application across the former sites compared with the latter provides evidence in favour of an account involving phonological vowel insertion, as the only plausible explanation of the difference is in terms of the frequency of ICIs blocking the process.

In this study, we investigated the application of voicing assimilation (henceforth ‘[VOICE] spreading’) in TLA plosive sequences. This has not been studied before, but work on other varieties of Arabic suggests that spreading of both voicing and voicelessness are commonly observed in Arabic consonant sequences (Al-Deaibes, 2016; Barry & Teifour, 1999; Kabrah, 2011; Teifour, 1997; Watson, 2002)—so this process seemed a good candidate for analysis in the present study. We reasoned that if we found evidence of the spreading of voicelessness across plosive hold phases (‘[-VOICE] spreading’) in our data, we could assess the evidence for some ICIs blocking this process. If we found evidence of the spreading of voicing (‘[+VOICE] spreading’), we would not expect epenthetic vowels to block this process, given their presumed underlying voicing. Epenthetic vowels might, however, contribute to [+VOICE] spreading patterns by themselves causing anticipatory or perseverative voicing in surrounding voiceless hold phases (see Barry & Teifour, 1999; Jansen, 2004).

If evidence for a phonological epenthesis rule in TLA is found, establishing the relative contributions of vocoid intrusion and vowel epenthesis to the observed data patterns requires a detailed comparison of ICIs across speakers and inter-consonantal sites. In particular, if epenthesis applies (near-) obligatorily in the inter-consonantal sites -CvC#C-, -Cv#CC-, and -CCv#CC-, consistent with Watson’s (2007) description, and optionally in other sites, we would expect to see no clear evidence of bimodality in ICI duration distributions in the former sites, and evidence of bimodality in the latter, due to the occurrence of both vocoid intrusion and low-frequency vowel epenthesis. Evidence of the long ICIs in these ‘other’ sites showing further characteristics of epenthetic vowels—for example, being fully voiced between voiceless consonants, or blocking [VOICE] spreading—would strengthen an analysis that includes optional epenthesis. By contrast, if all ‘optional vowel epenthesis’ in sites other than -CvC#C-, -Cv#CC-, and -CCv#CC- is really vocoid intrusion, we would expect the bimodality in ICI distributions that is observed across inter-consonantal sites to disappear once the data are split by site.

1.3. Factors influencing inter-consonantal timing

In contexts in which there is no clear evidence for a phonological epenthesis rule, we would expect the data patterns of ‘variable epenthesis’ to be accountable in terms of independently motivated factors influencing inter-consonantal timing, including at least the following.

First, it has been found repeatedly that onset clusters exhibit lower gestural overlap than coda clusters (Byrd, 1996; Chitoran, Goldstein, & Byrd, 2002; Gafos, 2002). Second, it has been found that consonant sequences across morpheme or word boundaries exhibit lower gestural overlap than clusters proper (Cho, 1998). Third, it has been found repeatedly that in stop clusters, ‘front-to-back’ clusters such as [tk] and [pt] exhibit greater gestural overlap than ‘back-to-front’ clusters such as [kt] and [tp] (Byrd, 1996; Chitoran et al., 2002; Surprenant & Goldstein, 1998; Zsiga, 1994). This means that it is reasonable to assume that the duration of intrusive vocoids varies systematically with ‘place order.’ It should be noted that Gafos et al. (2010) provide evidence for a ‘relativized place order hypothesis,’ which says that place order effects decrease in strength as ICI duration increases. This means that the absence of place order effects across relatively long ICIs should not be taken as straightforward evidence *against* an account in terms of vocoid intrusion.

Finally, Shaw and Davidson (2011) run simulations on computational models of gestural organization in plosive-plosive clusters, among other cluster types. They report on simulations with increasing random variability associated with the first plosive release, and on simulations with the same *and* low-frequency vowel insertion. The simulations with increasing random variability show a strong positive correlation between the mean and the standard deviation of ICI duration. This is consistent with the general property of motor behaviour that the variance of an interval is correlated with its mean (Schoner, 2002; Shaw, Gafos, Hoole, & Zeroual, 2009, 2011). The simulations that include vowel insertion show the same positive correlation at higher levels of variability. At lower levels of variability, however, the correlation is negative or non-existent. Although Shaw and Davidson (2011) do not offer an explicit account of this difference between the simulations, it is presumably due to the relative stability in the durations of the inserted vowels. The pattern suggests that while a positive correlation between the mean and standard deviation of ICI duration is an expected correlate of vocoid intrusion, its *absence* in specific contexts can be taken as evidence for phonological vowel insertion.

1.4. This study

As indicated above, this study was motivated by the report that TLA is characterized by widespread, partly optional vowel epenthesis in consonant strings (Watson, 2007). We assess the case for a phonological epenthesis rule by investigating the duration, voicing, and amplitude characteristics of inter-consonantal intervals in a range of plosive sequences produced by four TLA speakers. By taking into consideration factors known to constrain inter-consonantal timing, we also try to assess the contribution of variable open transition to TLA’s ‘variable epenthesis’ patterns. Moreover, we investigate plosive voicing to establish whether evidence can be found of specific inter-consonantal intervals blocking [VOICE] spreading.

Our methodological approach is in line with some of the work cited above on inter-consonantal coordination (e.g., Gafos et al., 2010) in relying on measurements of continuous phonetic parameters, and avoiding overt classifications of inter-consonantal intervals as consonant releases, intrusive vocoids, or epenthetic vowels. Most of the work on epenthesis cited by Hall (2006) relies on auditory transcriptions informed by analysts’ or native speakers’ judgements as to the status of particular vowels. Other studies rely on

overt classifications of ICIs by expert coders, based on criteria pertaining to continuous phonetic parameters: For example, in a study of non-native consonant cluster production reported by Wilson, Davidson, and Martin (2014), coders inspected spectrograms and waveforms of Russian target words produced by English speakers, and distinguished between first-consonant releases and (erroneous) vowel epenthesis in clusters by looking, among other things, for visible first and second formants. Hall (2013) discusses some of the challenges in eliciting reliable native-speaker judgements in this area of study, and Ridouane (2008) highlights the difficulty in delimiting devoiced vowels and preceding consonant releases. Wilson and Davidson (2015) show that expert coders' classification of Wilson et al. (2014) can be replicated by measurements taken across multiple acoustic parameters and processed using statistical classification techniques—circumventing the need for *a priori* statements of classification criteria. In this study we took a similarly agnostic approach to where the boundaries between releases, intrusive vocoids, and epenthetic vowels might be in perceptual terms, on the assumption that acoustic and distributional analyses of ICIs in plosive sequences can provide detailed insight into the nature of variable epenthesis in TLA.

2. Data and method

2.1. Data

We constructed a set of one and two-word test items that covers grammatical strings of two to four plosives within words and across word boundaries. (1) lists the string types and ICI sites covered by the data set. ‘##’ refers to an utterance boundary; ‘#’ to a word boundary. ICI sites are highlighted and numbered.

- (1)
- a. ##C¹C-
 - b. -C²C##
 - c. -C³#C-
 - d. -C⁴#C⁵C-
 - e. -C⁶C⁷#C-
 - f. -C⁸C⁹#C¹⁰C-

We designed the data set to analyze ICIs in the strings /t(#)k/, /k(#)t/, /g(#)d/, /t(#)g/, /g(#)t/, and /k(#)d/ at the epenthesis sites listed in (1). All of these strings occur both syllable-initially and syllable-finally as clusters, as well as occurring as sequences across word boundaries. We did not attempt to include items that would allow for systematic analysis of ICIs in /d(#)k/ and /d(#)g/ because the clusters /dk/ and /dg/ occur syllable-initially only. In identifying items covering the longer plosive sequences, we avoided items with adjacent homorganic plosives. We tried to maximize the use of /t/, /d/, /k/, and /g/ throughout the sequences, although we were constrained by our attempt to balance the data set for voicing (see below). In a small number of cases we used /b/ instead of /d/ or /g/ to construct a meaningful item. We also limited vowel qualities to /a/ and /a:/ as much as possible; in a small number of cases we used /e:/ to create a meaningful item. We maximized the use of individual test items: For example, we included /wagt#kdab/ ‘the time when he lied’ to yield three analysable ICIs: one in the word-final cluster /gt/ (-C³C#CC-), one at the word boundary /t#k/ (-CC⁹#CC-), and one in the word-initial cluster /kd/ (-CC#C¹⁰C-).

In order to be able to investigate [VOICE] spreading patterns, we ensured that all two-plosive strings embedded in longer sequences were surrounded by all logically possible combinations of (phonologically) voiced and voiceless plosives. To illustrate, for /t#k/ in -C⁴#CC- we included an item with a following voiced plosive (/ʁe:t#kdab/ ‘Gheet lied’)

and a corresponding item with a following voiceless plosive (/ha:t#kta:b/ ‘bring a book’). For /gt/ in -C¹C#CC-, we included four test items: one with a following voiceless cluster (/wagt#kta:b/ ‘time for a book’), one with a following voiced cluster (/wagt#bde:/ ‘time when he started’), and two with corresponding clusters with mixed voicing (/wagt#kdab/ ‘the time when he lied’ and /wagt#gta:/ ‘the time when he killed’).

These efforts yield a data set in which each ICI site listed in (1) above is represented by six items (sites 1, 2, and 3), 12 items (sites 4, 5, 6, and 7) or 24 items (sites 8, 9, and 10). We adjusted the balance by including two items for each string at sites 1, 2, and 3: this yields a total of 111 items in which each site listed in (1) above is represented by either 12 or 24 items. The full list of items is given in the appendix. The number of ICIs to analyze across the ten sites is $(6 \times 12 + 3 \times 24 =)$ 156. We elicited three productions of each item from four speakers (see below); the result is a total of $(4 \times 3 \times 156 =)$ 1872 ICIs to analyze. The number of plosive hold phases throughout the 111 items is 340. We excluded utterance-initial and final items for the purpose of hold phase analysis, leaving $(4 \times 3 \times 292 =)$ 3504 hold phases.

2.2. Participants and procedure

Four male speakers of TLA participated in this study; they are referred to hereafter as A, B, C, and D. All were born and raised in Tripoli and all use TLA as an everyday language. At the time of the recording, all participants were postgraduate students at the University of Leeds, ranging in age between 37 and 47. All provided explicit signed consent, and all were offered compensation for their time.

We produced three randomized lists of the 111 items, plus 30 fillers, and printed them in large Arabic script without vowel diacritics. Before the recording, the participants had the chance to practice producing the items and to annotate the lists with diacritics and short vowels, if they wished. They were not guided in this. Participants were asked to read items containing ICI site 1 followed by /halba/ ‘many times,’ and items containing ICI sites 2 to 10 preceded by /matgu:li:ʃ/ ‘don’t say (masculine).’ They were asked to speak at a normal speaking rate, and they were explicitly instructed to produce each item as they would in everyday, casual speech in Tripoli. The second author, who is a native speaker of TLA himself, supervised all recordings to ensure that the participants did this, as the use of Arabic script might bias them towards producing pronunciations influenced by Modern Standard Arabic. Where relevant, participants were reminded of the instruction and asked to repeat the last item; the reproduction was then recorded for analysis. Disfluencies were dealt with similarly. The recording for each speaker was done in a single session, with short breaks separating the productions of the three lists.

2.3. Acoustic parameters

Plosive hold phase and ICI durations were segmented in Praat (Boersma & Weenink, 2016) with reference to waveforms and wideband spectrograms. Hold phase duration was delimited between the sudden reduction in overall waveform amplitude at the end of a preceding vowel or ICI and the release burst, or where no burst was evident, the sudden increase in overall waveform amplitude at the start of the following vowel or ICI. ICI duration was delimited between the release of the first plosive and the hold phase onset of the second. The segmentation is illustrated in **Figure 1**. Since initial hold phases could not be reliably segmented in utterance-initial items (site 1, ##C¹C-) and final hold phases could not be reliably segmented in utterance-final items (site 2, -C²C##), we excluded these items from relevant analyses, as indicated above. The segmentation of ICIs revealed 11 instances of a zero ICI (out of 1872, <1%). In the relevant strings, the hold phases of the two plosives overlap, and there is no perceptible release of the first

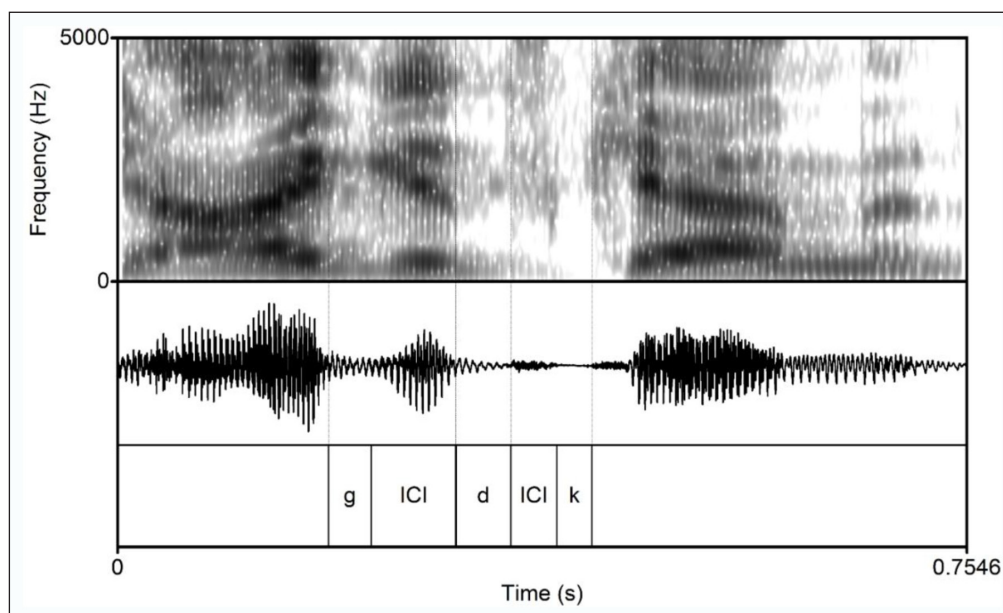


Figure 1: Waveform and spectrogram of /ʕagd#ka:mi/ 'Kamil's tying' (Speaker D, first repetition), with plosive hold phases and ICIs delimited.

plosive. We excluded these instances from our analyses, leaving a data set of 1861 ICIs. Durations were log-transformed for the purpose of quantitative analysis, although we will illustrate the main data patterns using the raw values. We will refer to the relevant variables as ICI Duration (ms), ICI Duration (log ms), Plosive Duration (ms), and Plosive Duration (log ms).

We used 'fraction of locally unvoiced frames' values extracted from voice reports in Praat to quantify surface voicing in ICIs and plosive hold phases (cf. Davidson, 2016, 2018). We used the pitch floor, pitch ceiling, and time step values recommended by Eager (2015) and subtracted each fraction from 1 to derive a 'fraction of locally voiced frames.' We will refer to the resulting variables as ICI Voicing Fraction and Plosive Voicing Fraction. For each segment, we extracted a voicing fraction for the entire segment, as well as fractions for successive thirds of the segment duration and fractions for the segment's first and second halves (cf. Davidson, 2016, 2018). Having established that in plosive hold phases, the middle third value is very rarely higher or lower than both the first and final third values (the 'hump' and 'trough' contours described by Davidson, accounting for < 5% of plosive hold phases), we decided that the fractions for hold phase halves were sufficiently informative for our purposes. We will refer to them in our analysis of [VOICE] spreading, where analyzing first-half and second-half voicing fractions may provide some insight into the direction of spreading patterns. For ICIs, we will use the fractions extracted for successive thirds to minimize the potential impact of plosive releases on our analysis of ICI voicing.

We also coded plosive hold phases for phonological, or 'underlying' voicing: That is, we coded /t/ and /k/ as voiceless and /d/, /g/, and /b/ as voiced. We will refer to the resulting variable as Plosive Underlying Voicing. Moreover, we derived the variable Relative Plosive Voicing Fraction by assigning underlyingly voiceless plosives a value of zero and underlyingly voiced plosives a value of 1, and then subtracting these values from Plosive Voicing Fraction values. On the resulting scale, a value of -1 marks an underlyingly voiced plosive which surfaces with no voicing; a value of zero marks a plosive whose surface voicing is consistent with its underlying status; and a value of 1 marks an underlyingly voiceless plosive which surfaces with full voicing.

We also extracted root mean square (RMS) amplitude values for all ICIs. As indicated above, previous studies (e.g., Wilson et al., 2014) have attempted to distinguish intrusive and epenthetic vowels by coding inter-consonantal intervals for the presence of visible formant structure above F2. In our data, this type of coding seems only partly informative. The first delimited ICI in **Figure 1** clearly has visible formant structure akin to that of the surrounding lexical vowels; the second delimited ICI is voiceless. The delimited ICI in **Figure 2** (top), /hag#da:mi/, is voiced, but does not have visible formant structure. This ICI is hearable as a voiced release of /g/. Both delimited ICIs in **Figure 2** (bottom), /hmad#kdab/, are voiced. Both also have visible formant structure, and both are hearable as a voiced release followed by a vowel-like segment. However, the formants are considerably more prominent in the first ICI than in the second, and the percept of

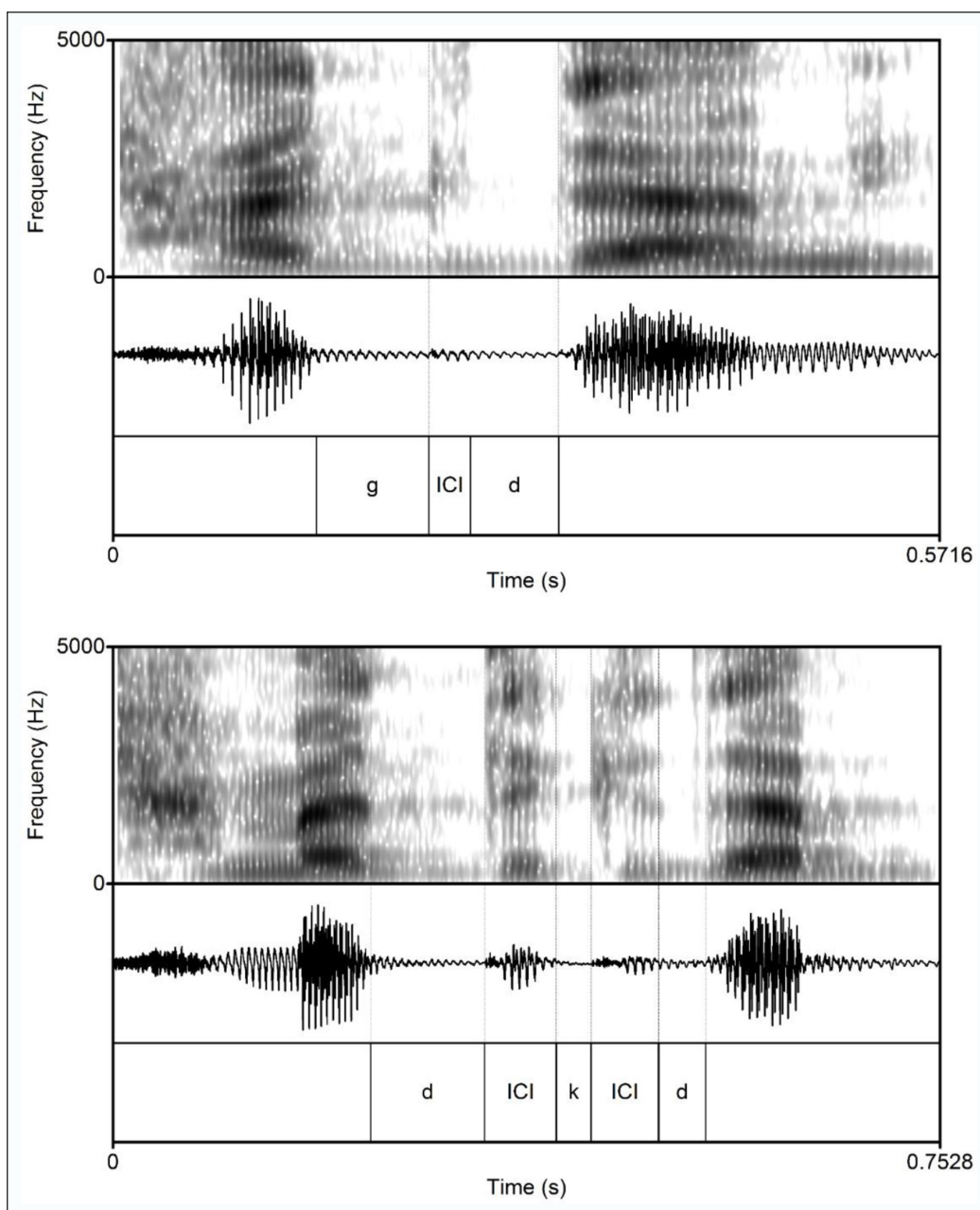


Figure 2: Waveforms and spectrograms of (top) /hag#da:mi/ ‘unbelievable right’ (Speaker D, first repetition) and (bottom) /hmad#kdab/ ‘Hmad lied’ (Speaker D, first repetition) with plosive hold phases and ICIs delimited.

a vowel is stronger—probably also due to the first ICI’s greater duration. Binary coding would not capture this variation, and setting a threshold for ‘minimum visibility’ would not be straightforward. A measure of overall acoustic intensity such as RMS amplitude (e.g., Wayland & Jongman, 2003) does capture some of this variation, on a continuous scale: The higher the amplitude measured across a voiced ICI, the more likely it is that the ICI is hearable as a full vowel. We will refer to the resulting variable as ICI RMS Amplitude. After extracting RMS amplitude values, we inspected the waveforms of ICIs with positive outlier values. Where it seemed clear that the RMS value was due to signal perturbations independent of the speaker’s ICI production, we excluded the ICI from our analyses. We excluded a total of 35 ICIs (< 2%), leaving a core data set of 1826 ICIs.

2.4. Phonological parameters

We coded each ICI for its site of occurrence (1 to 10); we will refer to the resulting variable as ICI Site. We used additional codings to group sites by shared phonological characteristics. First, we distinguished sites in which (near-)obligatory epenthesis might be expected from the rest. We will refer to the resulting variable as Predicted Epenthesis. Second, we distinguished ICI sites in onset, coda, and word boundary positions. We will refer to the resulting variable as Syllabic Context. **Table 1** cross-tabulates the levels of these three variables. Finally, we coded each ICI for occurring in a *front-to-back* plosive string (/t(#)k/, /t(#)g/) or in a *back-to-front* one (/k(#)t/, /k(#)d/, /g(#)d/, /g(#)t/). In what follows, we will refer to the resulting variable as Place Order, with the levels ‘FB’ and ‘BF.’

2.5. Quantitative methods

All quantitative analysis was performed in R (R Core Team, 2016). The main packages we used were *lme4* (Bates, Maechler, Bolker, & Walker, 2015) for linear mixed effects modelling (version 1.1-12), *cluster* (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2016) for cluster analysis, and *partykit* (Hothorn & Zeileis, 2015) and *randomForest* (Liaw & Wiener, 2002) for analysis using regression trees. We explain these methods further as we describe the analysis below. For reference, **Table 2** lists the variables introduced so far, including categorical variable levels. We will introduce several derived variables in the course of the analysis below.

Table 1: Cross-tabulation of levels for ICI Site, Predicted Epenthesis, and Syllabic Context.

ICI Site	Predicted Epenthesis	Syllabic Context
1 (##C①C-)	no	onset
2 (-C②C##)	no	coda
3 (-C③#C-)	no	boundary
4 (-C④#CC-)	yes	boundary
5 (-C#C⑤C-)	no	onset
6 (-C⑥C#C-)	yes	coda
7 (-CC⑦#C-)	no	boundary
8 (-C⑧C#CC)	no	coda
9 (-CC⑨#CC-)	yes	boundary
10 (-CC#C⑩C-)	no	onset

Table 2: Main continuous (acoustic) and categorical (phonological) variables, with levels for the latter.

Variable type	ICIs	Plosives
Continuous	ICI Duration (ms, log ms)	Plosive Duration (ms, log ms)
	ICI Voicing Fraction	Plosive Voicing Fraction
	ICI RMS Amplitude	Relative Plosive Voicing Fraction
Categorical	ICI Site ('1'–'10')	Plosive Underlying Voicing ('voiced,' 'voiceless')
	Predicted Epenthesis ('yes,' 'no')	
	Syllabic Context ('onset,' 'coda,' 'boundary')	
	Place Order ('FB,' 'BF')	

3. Results

In what follows, we first describe the distributions of our main acoustic parameters (3.1). We address the extent to which multimodality can be observed, compare the distributions of our four speakers, and show what the distributions look like when split by phonological environment. We then report on a cluster analysis (3.2) aimed at identifying distinct ICI 'shapes' across the acoustic parameters; these shapes can then be mapped to phonological variables. We go on to present separate models for ICI voicing (3.3) and ICI duration (3.4), which allow us to address multiple predictions of alternative accounts of variable epenthesis specific to these two parameters. Finally, we report on an analysis of plosive hold phase voicing (3.5) aimed at establishing whether ICIs of different acoustic 'shapes' and in different phonological environments interact with [VOICE] spreading patterns.

3.1. Descriptives

We begin by inspecting the distributions of ICI Voicing Fraction, ICI Duration (ms), and ICI RMS Amplitude. **Figure 3** shows density kernel plots for each split by Speaker. In relation to ICI Voicing Fraction, an important question is whether voiced ICIs do indeed occur in our data. This is clearly the case. The four speakers' distributions are very similar in shape, showing clear multimodality: They include a subset of values close or equal to zero (likely to be perceived as fully voiceless), values close or equal to 1 (likely to be perceived as fully voiced), and intermediate values representing partial voicing. All four distributions are skewed towards values above 0.5: That is, ICI voicing is relatively common.

In relation to ICI Duration (ms), as indicated above, if TLA 'variable epenthesis' derives from phonological vowel insertion as well as variable open transition, we would expect to find some evidence of bimodality in duration distributions (cf. Bellik, 2018). *Some* evidence of bimodality can indeed be observed. The four speakers all show a main peak around 25 ms, and either a second peak or a 'right shoulder' in the range 50–75 ms. That is, all four speakers produce a majority of ICIs with durations around 25 ms, and a smaller subset of instances with durations around or above 50 ms. The second peak is most distinct for speaker C. Hartigan's dip tests (Maechler, 2015) for the (log-transformed) distributions show that C's distribution is significantly different from unimodal ($D = 0.026$, $p = 0.028$); the other speakers' distributions are not significantly different from unimodal. Similarly, we might expect to see some evidence of multimodality in the ICI RMS Amplitude distributions, on the assumption that ICIs containing voiced epenthetic vowels will have considerably higher amplitudes than other ICIs. While **Figure 3** shows some evidence of a 'right tail' of ICI RMS Amplitude values for all four speakers, this is not enough to yield significant divergence from unimodality according to Hartigan's dip tests. We do see considerable inter-speaker variation: B and D have wider ranges of values than A and C, and more even spreads across those ranges.

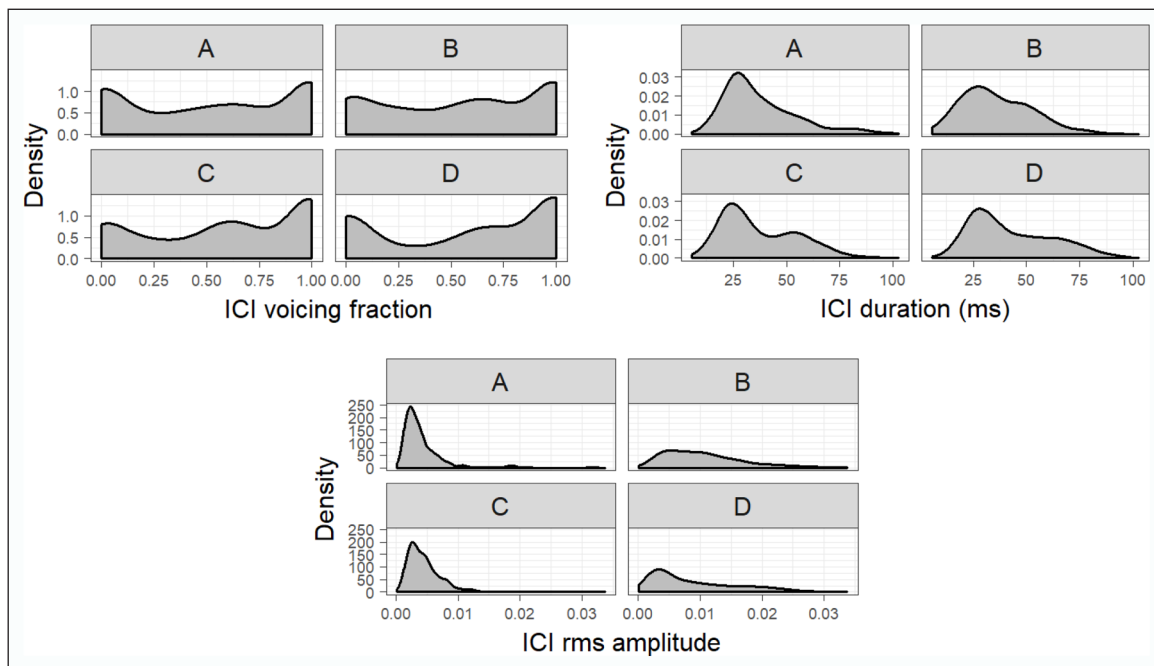


Figure 3: Kernel distribution plots for ICI Voicing Fraction (top left), ICI Duration (ms) (top right), and ICI RMS Amplitude (bottom), split by Speaker.

Before moving to more detailed modelling of ICI voicing and duration, we probe the relationship between ICI voicing, duration, and amplitude on the one hand, and inter-consonantal site on the other. Given Watson's (2007) account of variable epenthesis, we can predict that ICI sites 4, 6, and 9 (-C⁴#CC-, -C⁶C#C-, -CC⁹#CC-), in which phonological vowel insertion is phonotactically motivated, to have different phonetic 'profiles' from ICI sites 1, 3, 5, 7, 8, and 10 (##C¹C-, -C³#C-, -C#C⁵C-, -CC⁷#C-, -C⁸C#C¹⁰C-), in which vowel insertion is either less frequent or absent. If it is rare in the latter contexts, we can further predict that the bimodality observed for ICI Duration (ms) will disappear when the data are viewed by ICI Site (see Torreira and Ernestus, 2011 for similar reasoning).

Figure 4 shows violin plots for each of ICI Voicing Fraction, ICI Duration (ms), and ICI RMS Amplitude. Violin plots are alternatives to boxplots in which each shape is effectively a kernel distribution plotted vertically and mirrored to highlight shape differences; the fatter the shape, the greater the density. For comparison against Watson's account, violins for the three predicted epenthesis sites (4, 6, 9) are set off against the others. The plots show that each of the three parameters is conditioned by ICI Site, and the conditioning is consistent across parameters—but the conditioning is not fully consistent with our predictions. ICI sites 2, 4, 6, and 9 are associated with different distribution shapes compared with ICI sites 1, 3, 5, 7, 8, and 10. To be more precise, sites 2, 4, 6, and 9 are associated with ICI Voicing Fraction values above 0.5, while sites 1, 3, 5, 7, 8, and 10 comprise sizeable subsets of instances with values below 0.25; they are associated with ICI duration distributions centring around or above 50 ms, while those at sites 1, 3, 5, 7, 8, and 10 are centred around 25 ms; and they are associated with considerably wider dispersals of ICI RMS Amplitude values. With reference to ICI duration, Hartigan's dip tests confirm that none of the ICI Duration (log ms) distributions by ICI Site are significantly different from unimodal.

3.2. Cluster analysis

To assess how the phonological conditioning illustrated in **Figure 4** constrains the correlations among the three phonetic parameters, we performed a cluster analysis on the variables ICI Voicing Fraction, ICI Duration (log ms), and ICI RMS Amplitude. To

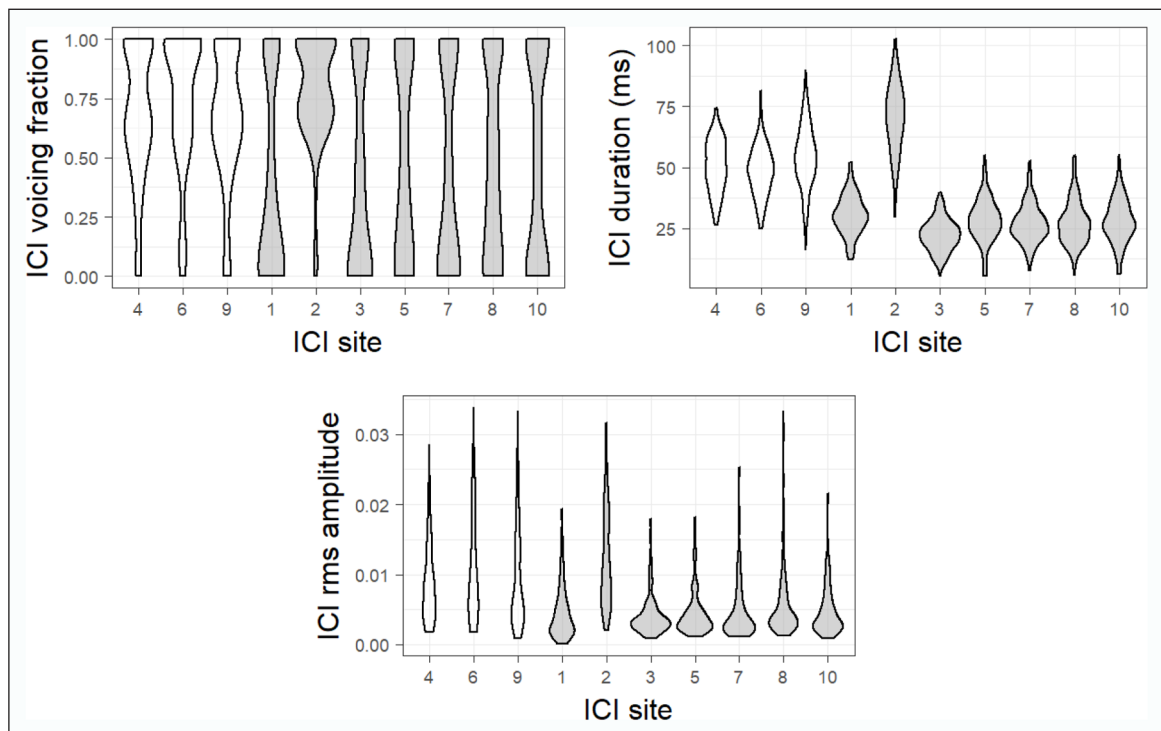


Figure 4: Violin plots for ICI Voicing Fraction (top left), ICI Duration (ms) (top right), and ICI RMS Amplitude (bottom), split by ICI Site; in each, violins for the three predicted epenthesis sites (4, 6, 9) are in white on the left.

restrict the influence of speaker differences on the clustering, we replaced each variable by the residuals of a linear mixed effects model of its distribution with Speaker as a random intercept. As our dataset is small and we are mostly interested in top-level clustering, we chose the divisive hierarchical analysis method (see Kaufman & Rousseeuw, 1990), implemented with default settings. Unlike k-means clustering, this method does not require the user to pre-specify a number of output clusters; rather, it generates a dendrogram which can then be inspected and ‘cut’ to yield different numbers of output clusters, with reference to diagnostics quantifying the relationship between inter and intra-cluster variability for different cluster numbers. In the case of our analysis, cutting to three clusters was optimal in terms of ‘average silhouette’; this yielded a three-level variable which we will call Cluster.

Figure 5 shows scatterplots illustrating the inter-correlations among ICI Voicing Fraction, ICI Duration (log ms), and ICI RMS Amplitude, split by Cluster. Cluster 1 comprises ICIs with a full range of ICI Voicing Fraction values (0–1), relatively low ICI Duration (log ms) values (mostly < 1.75), and relatively low ICI RMS Amplitude values (mostly < 0.01). Cluster 2 comprises relatively high ICI Voicing Fraction values (mostly > 0.35), relatively high ICI Duration (log ms) values (mostly < 1.5), and ICI RMS Amplitude values centred slightly higher than in cluster 1 (≈ 0.08). Cluster 3 comprises relatively high ICI Voicing Fraction and ICI Duration (log ms) values, as in cluster 2, combined with relatively high ICI RMS Amplitude values (< 0.01).

We then fitted a conditional inference regression tree model (Baayen, 2013; Strobl, Malley, & Tutz, 2009; Tagliamonte & Baayen, 2012) for Cluster, with ICI Site and Speaker as conditioning variables. The resulting tree is shown in **Figure 6**. The tree results from a recursive binary partitioning algorithm. Given a dependent, or ‘response’ variable and a set of conditioning variables, the algorithm establishes for each conditioning variable whether ‘splitting’ the data on a particular value results in two homogeneous subsets. If this is the case for multiple conditioning variables, the algorithm implements the split

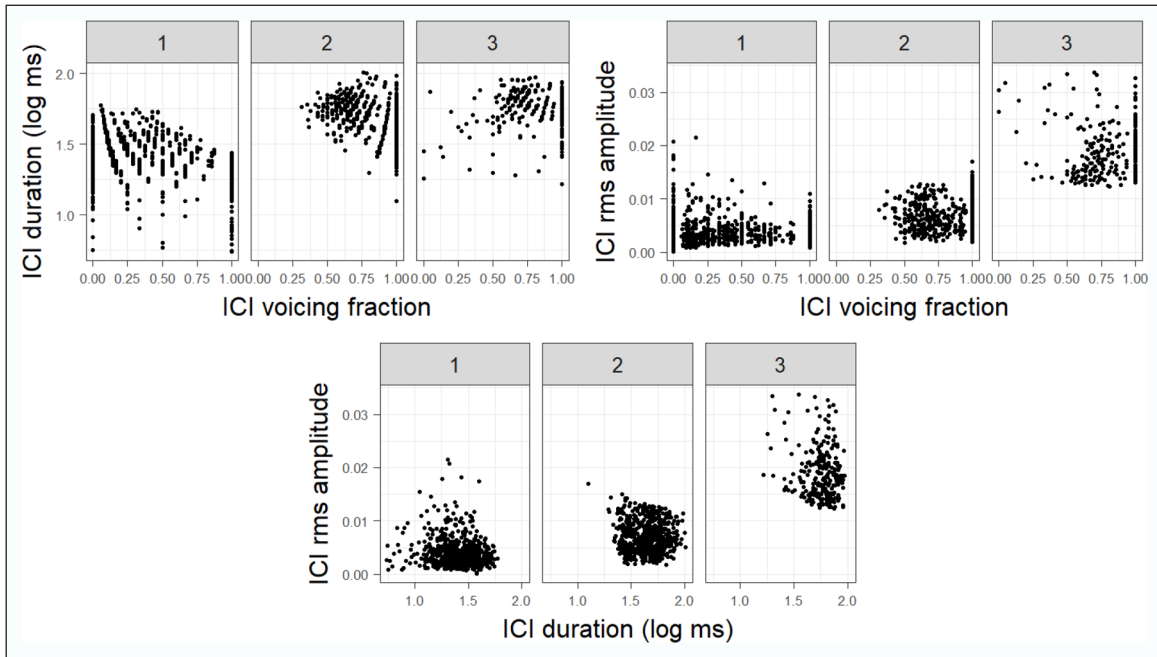


Figure 5: Scatter plots for pairwise correlations between ICI Voicing fraction, ICI Duration (log ms), and ICI RMS Amplitude, split by Cluster.

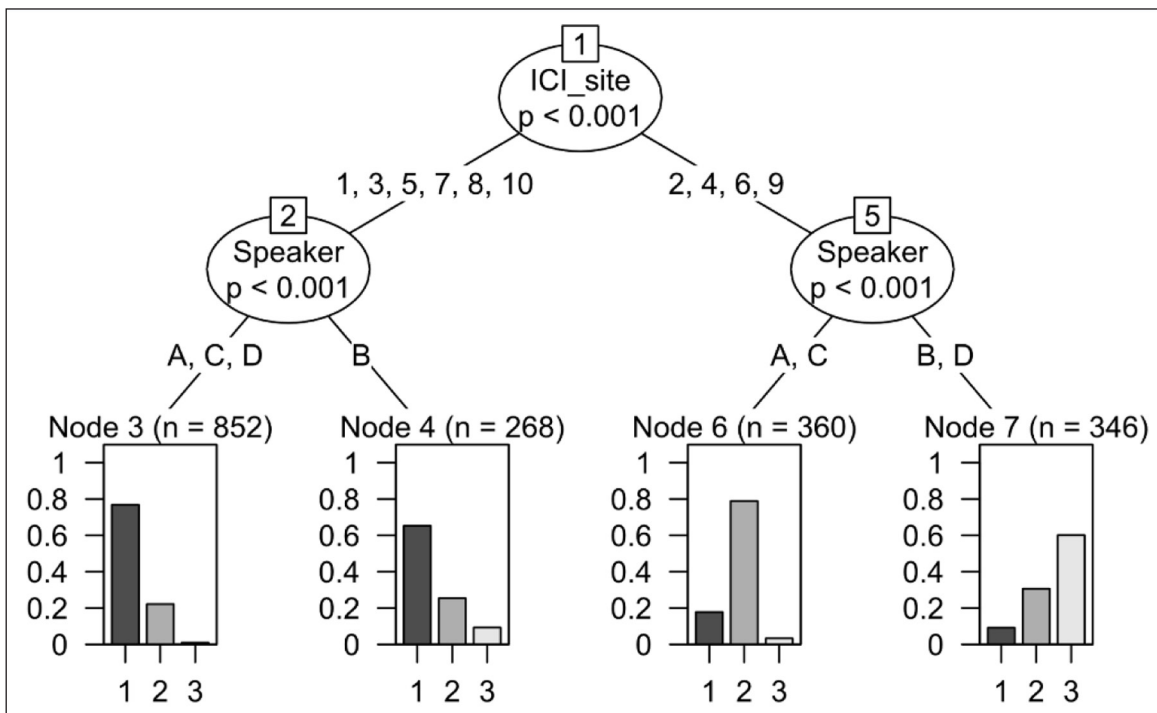


Figure 6: Conditional inference regression tree for Cluster with ICI Site and Speaker as conditioning variables; the bar plots at the terminal nodes show proportions of ICIs in cluster 1, 2, and 3.

associated with the greatest homogeneity; this is shown as the top node (numbered 1) in the resulting tree model. It then repeats itself within each of the two resulting subsets of data. It keeps implementing splits, adding further binary nodes in the tree model, until no further homogeneous subsets can be created. In the resulting tree model, the ‘terminal nodes’ are associated with plots showing the behaviour of the response variable in the

relevant smallest subsets of data, and the size of these subsets; to interpret the splits higher up, one must generalize across these plots. In the case of **Figure 6**, the terminal nodes (numbered 3, 4, 6, and 7) show the proportions of ICIs in cluster 1, 2, and 3. (See Plug & Carter, 2013, 2014; Strycharczuk, Van 't Veer, Bruil, & Linke, 2014 for further examples of the use of this modelling technique in laboratory phonology studies.)

Figure 6 confirms that the phonological conditioning seen in **Figure 4** is systematic both across the three phonetic parameters and across speakers: ICI Site yields the first, main split in the data, separating sites 1, 3, 5, 7, 8, and 10 from sites 2, 4, 6, and 9. Looking at the bar plots at the terminal nodes, we see that sites 1, 3, 5, 7, 8, and 10 are characterized by high proportions of cluster 1 (variable voicing, low duration, low amplitude), while sites 2, 4, 6, and 9 are characterized by high proportions of clusters 2 and 3 (high voicing, high duration, mid or high amplitude). Despite taking speaker differences into account in constructing the cluster model, Speaker yields significant splits 'underneath' ICI Site which closely mirror the grouping of speakers we have seen in **Figure 3**. Looking again at the bar plots at the terminal nodes, we see that speaker B is different from A, C, and D in having some cluster 3 ICIs in sites 1, 3, 5, 7, 8, and 10 (node 3 versus node 4), and B and D are both different from A and C in having mostly cluster 3 ICIs (high voicing, high duration, high amplitude) at sites 2, 4, 6, and 9; A and C show a clear preference from cluster 2 ICIs (high voicing, high duration, mid amplitude) in these sites (node 6 versus node 7).

Recall the prediction that ICI sites 4, 6, and 9, in which phonological vowel insertion is phonotactically motivated, have different phonetic 'profiles' from ICI sites 1, 2, 3, 5, 7, 8, and 10, in which vowel insertion should either be less frequent or absent altogether. The results shown so far suggest that this prediction is false for our data—not because sites 4, 6, and 9 do not stand out, but because site 2 (-C $\text{\textcircled{2}}$ C##) clusters with them. As this is not a site in which epenthesis is expected on phonotactic grounds, replacing ICI Site with a binary factor coding for phonotactically motivated epenthesis (Predicted Epenthesis) results in a poorer model of the data. Similarly, it is difficult to see how the observed grouping of sites could be seen as falling out of general factors influencing inter-consonantal timing. In particular, sites 2, 4, 6, and 9 include no onset ICIs (which one might predict to be longer than coda ICIs), but both coda and word boundary ICIs (of which one might predict the latter to be longer). As seen above, the site with the highest mean ICI duration is in fact site 2, which is a coda site; moreover, sites 4 and 6 show similar distributions despite the former occurring at a word boundary and the latter in a coda. This means that Predicted Epenthesis and a factor separating onset, coda, and word boundary ICIs (Syllabic Context) still do not capture the variance in the data as well as ICI Site.

This is confirmed by a random forest model (Baayen, 2013; Tagliamonte & Baayen, 2012; Tomaschek, Hendrix, & Baayen, 2018) for Cluster, including the conditioning variables Speaker, ICI Site, Predicted Epenthesis, and Syllabic Context. A random forest is a collection of regression tree models, each fit on a subset of the data under investigation, and with a subset of the specified conditioning variables. This allows for a robust assessment of the relative importance of individual conditioning variables, even when these are inter-correlated: See Plug and Carter (2014) for an illustration of this. Our forest model comprised 500 trees. Inspection of gini values, which quantify model fit across trees, confirms that ICI Site is by far the most informative conditioning variable (mean decrease = 193), followed by Speaker (mean decrease = 104), and, at considerable distance, Predicted Epenthesis (mean decrease = 51) and Syllabic Context (mean decrease = 19). Rerunning the model without ICI Site projects Predicted Epenthesis to first place

in terms of variable importance; however, its mean gini decrease value is considerably lower than that of ICI Site in the first model (118 vs 193); in other words, the best model fit is attained by including ICI Site.

3.3. ICI voicing

3.3.1. Preliminaries

Having established that two groups of ICI sites are associated with distinct phonetic ‘profiles,’ we return to our analysis of the individual phonetic parameters of ICI voicing and ICI duration, in order to probe parameter-specific data patterns that might narrow down the range of reasonable accounts for ‘variable epenthesis’ in TLA. Starting with ICI voicing, the occurrence of voiced ICIs between voiceless plosives can be taken as strong evidence in favour of phonological vowel insertion, as there is no obvious contextual source for the observed voicing. Given Watson’s (2007) account, we would expect that if such ICIs are present in our data, it will be in inter-consonantal sites 4, 6, and 9; we have seen above that site 2 patterns with these sites, but plosive voicing was not analyzed in sites 1 or 2. So far we have seen that sites 4, 6, and 9 are characterized by the near-absence of voicing fractions below 0.5, which are common elsewhere.

3.3.2. Analysis

Figure 7 shows ICI Voicing Fraction values plotted against values derived by averaging across the voicing fractions of the immediately preceding and following plosive hold phases. The plotting is done by ICI Site, and reflects density: Light grey dots correspond to single data points, and darker areas in the scatter indicate that there are many individual data points with similar values. For the purpose of this analysis we calculated ICI Voicing Fraction not across the entire ICI, as above, but across the middle and last thirds only. We did this because our ICIs include plosive releases, so that a fully voiceless plosive followed

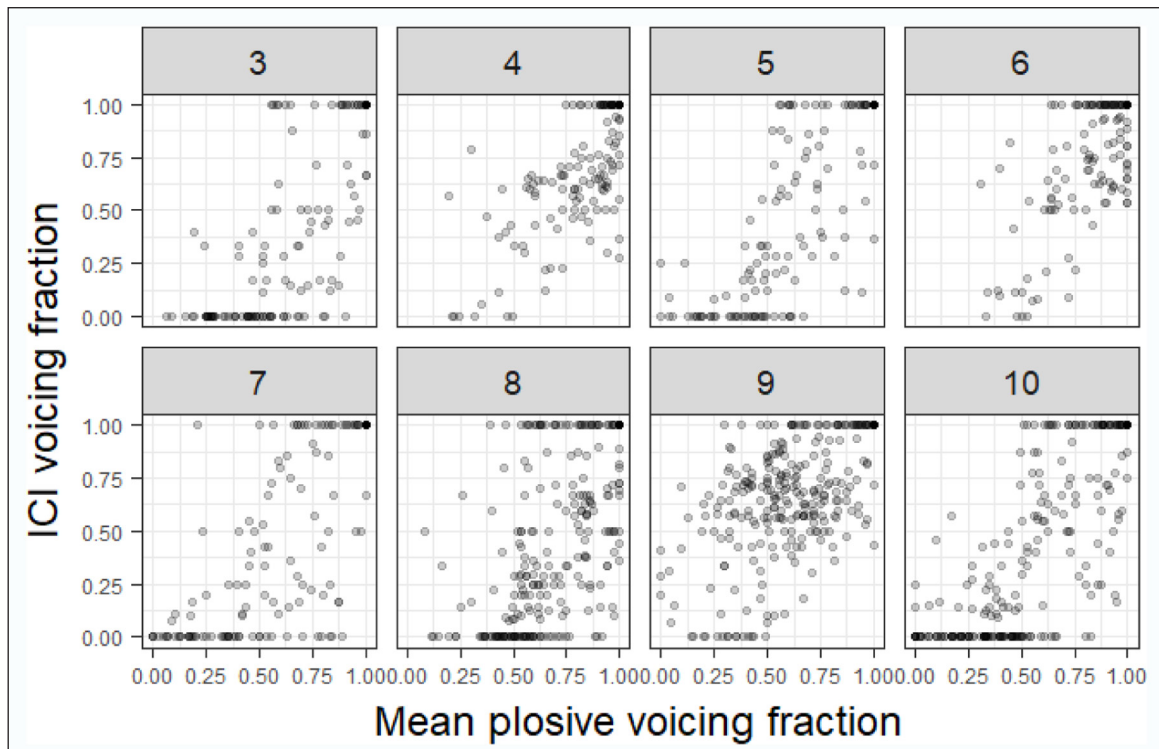


Figure 7: Density scatter plot for ICI Voicing Fraction (calculated excluding the first third of the ICI) and Mean Plosive Voicing Fraction, split by ICI Site.

by a voiced vowel would necessarily result in an ICI voicing fraction below 1. Excluding the first third of the ICI should exclude most plosive releases, particularly in longer ICIs. The scatter suggests that our data set contains no fully voiced ICIs—excluding most plosive releases—between two fully voiceless hold phases (y-axis 1, x-axis 0), and there are very few fully or near-fully voiced ICIs between hold phases whose voicing fractions average below 0.5. Fully voiceless ICIs co-occur with a wider range of hold phase voicing fractions, mostly centred below or around 0.5. There are no fully voiceless ICIs between two fully voiced hold phases (y-axis 0, x-axis 1). In other words, there is some evidence here for conditioning of ICI voicing by the voicing of the surrounding hold phases. This conditioning appears to vary between sites: Sites 4, 6, and 9 have few (near-)voiceless ICIs and few low-voicing ICIs between low-voicing hold phases.

In modelling ICI Voicing Fraction (again excluding the first third of the ICIs, although modelling voicing fractions across the entire ICIs leads to the same conclusions), we first confirmed that ICI Site is a significant predictor. Adding this to a linear mixed effects model with a significant random intercept for Speaker significantly improves fit ($\chi^2 = 250$, $df = 7$, $p < 0.001$). (A random effect for item repetition did not improve model fit, so was left aside.) Mean Plosive Voicing Fraction improves model fit further, particularly when added as an interaction with ICI Site ($\chi^2 = 1063$, $df = 8$, $p < 0.001$). The interaction is visualized in the conditional inference regression tree in **Figure 8**. Here ICI Site is replaced by Predicted Epenthesis, which, with sites 1 and 2 excluded from the analysis, separates sites 4, 6, 9 (‘yes’) from sites 3, 5, 7, 8, and 10 (‘no’). We checked whether residualizing ICI voicing fraction for Speaker made a difference to the tree model; it did not, so we present the tree with raw voicing fraction values on the y-axes of the terminal box plots. The tree model confirms that when the voicing fraction across the two hold phases is relatively low (up to 0.55), ICIs in sites 4, 6, and 9 have higher voicing fractions

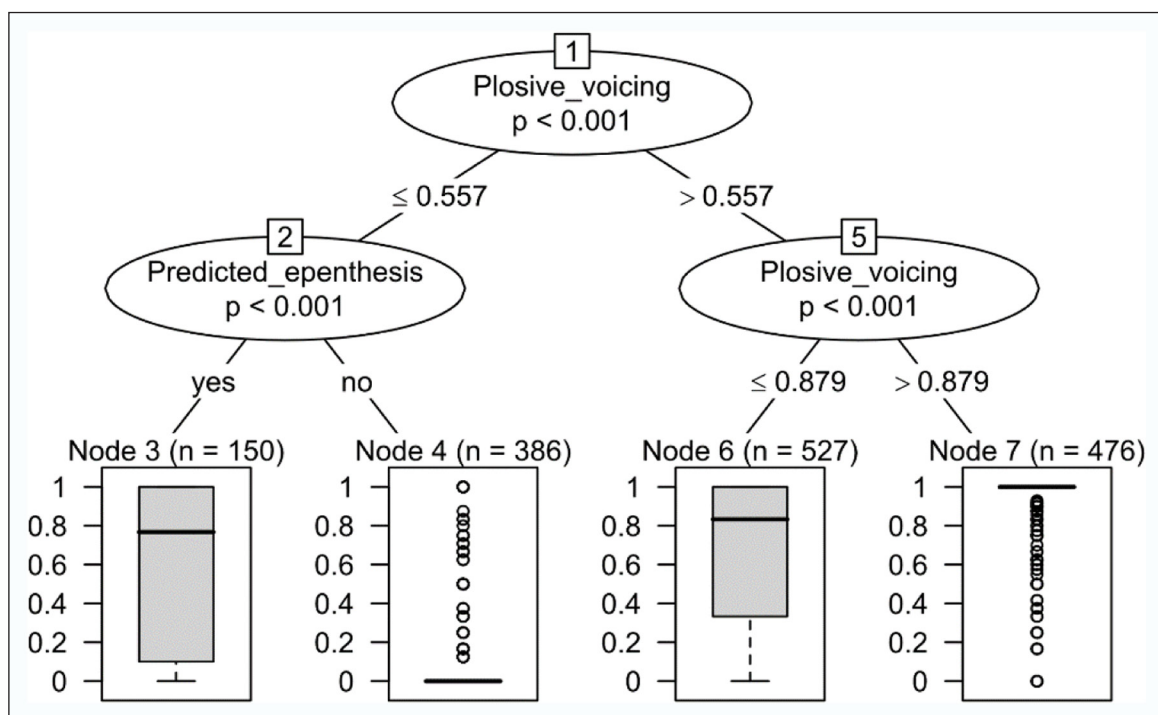


Figure 8: Conditional inference regression tree for ICI Voicing Fraction (calculated excluding the first third of the ICI) with Predicted Epenthesis and Mean Plosive Voicing Fraction (abbreviated here to Plosive Voicing) as conditioning variables; the box plots at the terminal nodes show the distributions of ICI Voicing Fraction in the relevant subsets of data.

than ICIs in sites 3, 5, 7, 8, and 10 (node 2): The former have a median voicing fraction around 0.8 (see the boxplot at node 3), while the latter are mostly (near-)voiceless (see the boxplot at node 4). This means that while ICI sites 4, 6, and 9 are not distinguished by the occurrence of fully voiced ICIs between voiceless hold phases, they are associated with significantly more substantial voicing in contexts with relatively little voicing. We will return to this finding in our [VOICE] spreading analysis below.

3.4. ICI duration

3.4.1. Preliminaries

In modelling ICI Duration (log ms), we focused our attention specifically on the extent to which ICI duration is conditioned by, or co-varies with, relevant characteristics of the surrounding plosives. As indicated above, it has been found repeatedly that ‘front-to-back’ stop clusters exhibit greater gestural overlap than ‘back-to-front’ ones: That is, ‘back-to-front’ clusters exhibit longer ICIs. It has also been observed that while epenthetic vowels add significantly to the duration of the consonant cluster they appear in, intrusive vowels do not. If the long ICI durations in sites 4, 6, and 9 are at least partly due to phonological vowel insertion, and the shorter durations in sites 3, 5, 7, 8, and 10 are primarily due to variable open transition, one might expect place order effects to be observed in the latter sites, not the former—or at least more robustly in the ‘open transition’ sites. Moreover, we would expect these two sets of sites to show different correlations between ICI durations and the durations of the surrounding hold phases: While this correlation should be negative for intrusive vowels, as total cluster duration is relatively stable, it should not be negative for epenthetic vowels, and might be expected to be positive on the assumption that both hold phase and epenthetic vowel duration are constrained similarly by the speaker’s articulation rate.¹

3.4.2. Analysis

We first confirmed that ICI Site is a significant predictor: Adding this to a linear mixed effects model with a random intercept for Speaker significantly improves fit ($\chi^2 = 1725$, $df = 9$, $p < 0.001$). (A random intercept for item repetition did not improve model fit, so was left aside.) We then assessed the relevance of Place Order and Mean Plosive Duration (log ms), which is calculated for each ICI across the two hold phases surrounding it. Expanding the model with an interaction between ICI Site and Place Order ($\chi^2 = 140$, $df = 10$, $p < 0.001$) results in a greater improvement of model fit than expanding it with a main effect for Place Order (AIC = -2401 versus -2335). Before assessing the relevance of Mean Plosive Duration (log ms), we refitted the model excluding ICI sites 1 and 2 (AIC = -1938). Again, expanding the model further with an interaction between ICI Site and Mean Plosive Duration (log ms) ($\chi^2 = 194$, $df = 15$, $p < 0.001$) results in a greater improvement of model fit than expanding it with a main effect for Mean Plosive Duration (log ms) (AIC = -2042 versus -2009). However, level comparisons suggest that these significant interactions do not straightforwardly reflect the hypothesized data patterns sketched above, which set apart sites 4, 6, and 9: rather, they are due to more

¹ We should note that Shaw and Kawahara (2018) make the opposite prediction in an analysis of C_1vC_2 sequences in Tokyo Japanese, in which v is a ‘devoiced vowel’ which may or may not be associated with a dedicated lingual gesture. They argue that under ‘CV coordination,’ “shortening of C_1 would expose more of the inter-consonantal interval,” while under ‘CC coordination,’ the release of C_1 and the onset of C_2 are in a stable temporal relationship. It is difficult to reconcile this with the finding that intrusive vocoids do not add to $C(v)C$ duration (Ridouane & Fougeron, 2011). As seen below, on the logic of Shaw and Kawahara (2018) we might have to conclude that sites 1, 3, 5, 8, and 10 are epenthesis sites and sites 4 and 6 are not, as judging by the correlations between ICI and hold phase durations, the former would appear to display CV coordination, and the latter CC coordination. We do not take this point further.

fine-grained variation between inter-consonantal sites in the strength of place order effects and covariance with plosive hold phase durations. Modelling ICI Duration (log ms) with conditional inference regression trees confirms this variation and allows for straightforward visualization.

The interaction between ICI Site and Place Order is illustrated in the tree in **Figure 9**, which visualizes a model of ICI Duration (log ms) with ICI Site and Place Order as conditioning variables; Place Order has the levels ‘BF’ (back-to-front) and ‘FB’ (front-to-back). The splits for Place Order reveal a consistent place order effect across most sites, including all sites in which epenthesis is expected on phonotactic grounds (that is, sites 4, 6, and 9). However, in two sites no significant place order effect is observed (see nodes 2 and 7): site 2, which is associated with the highest mean ICI duration, and site 3, which is associated with the lowest.

The interaction between ICI Site and Mean Plosive Duration (log ms) is illustrated in the tree in **Figure 10**, which visualizes a model of ICI Duration (log ms) with ICI Site and Mean Plosive Duration (log ms) as conditioning variables. The tree shows that Mean Plosive Duration (log ms) has a consistent effect across sites 3, 5, 7, 8, and 10 (see node 7), such that lower plosive durations are associated with higher ICI durations. This effect is also observed for site 9 (-CC⁹#CC-, see node 3), but it is not observed for sites 4 and 6 (-C⁴#CC-, -C⁶C#C-). A simple correlation analysis on ICI Duration (log ms) and Mean Plosive Duration (log ms) (each residualized for Speaker) confirms that site 9 is associated with a negative correlation whose specifics ($r = -0.18$, 95% confidence interval upper limit -0.08) are very similar to those associated with sites at which epenthesis is not predicted to be widespread (for example, for site 8 $r = -0.18$, 95% confidence interval upper limit -0.06). Site 4 is associated with a positive correlation in this analysis ($r = 0.41$, 95% confidence interval lower limit 0.26), and site 6 with no correlation ($r = 0.00$, 95% confidence interval $-0.16-0.17$).

In sum, we find no support for the hypothesis that place order effects are weaker for ICIs in predicted epenthesis sites; in fact, a consistent place order effect is observed across

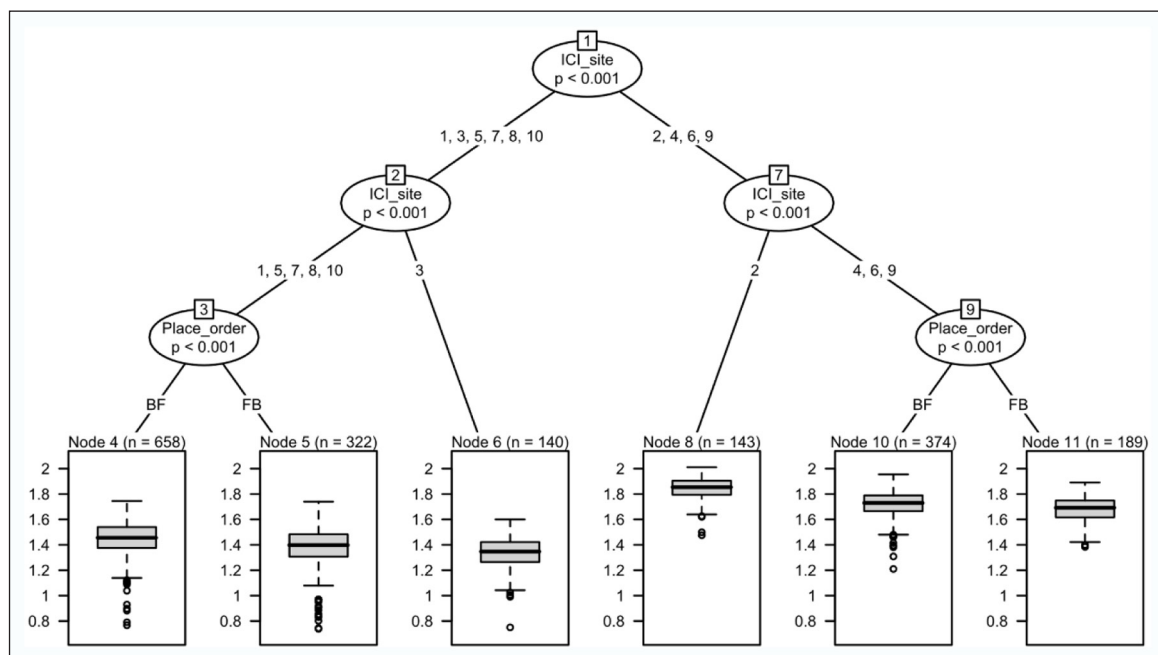


Figure 9: Conditional inference regression tree for ICI Duration (log ms) with ICI Site and Place Order as conditioning variables; the box plots at the terminal nodes show the distributions of ICI Duration (log ms) in the relevant subsets of data.

most sites, including sites 4, 6, and 9. We do find some support for the hypothesis that ICIs at predicted epenthesis sites are in a different temporal relationship with the surrounding plosive hold phases: While there is evidence of a negative relationship in site 9, along with all sites in which epenthesis is *not* expected, ICIs at the other predicted epenthesis sites appear temporally independent from the surrounding hold phases, or positively related.

3.4.3. A note on ICI duration variance

Finally, we noted above that the simulations run by Shaw and Davidson (2011) suggest that the absence of positive correlations between the means and standard deviations of ICI durations across environments can be taken as evidence for the occurrence of phonological vowel insertion. Inspection of ICI duration means and standard deviations by ICI Site shows that raw durations yield strong positive correlations for all four speakers. The pattern is illustrated in **Figure 11** (across speakers, Pearson's $r = 0.63$, 95% confidence

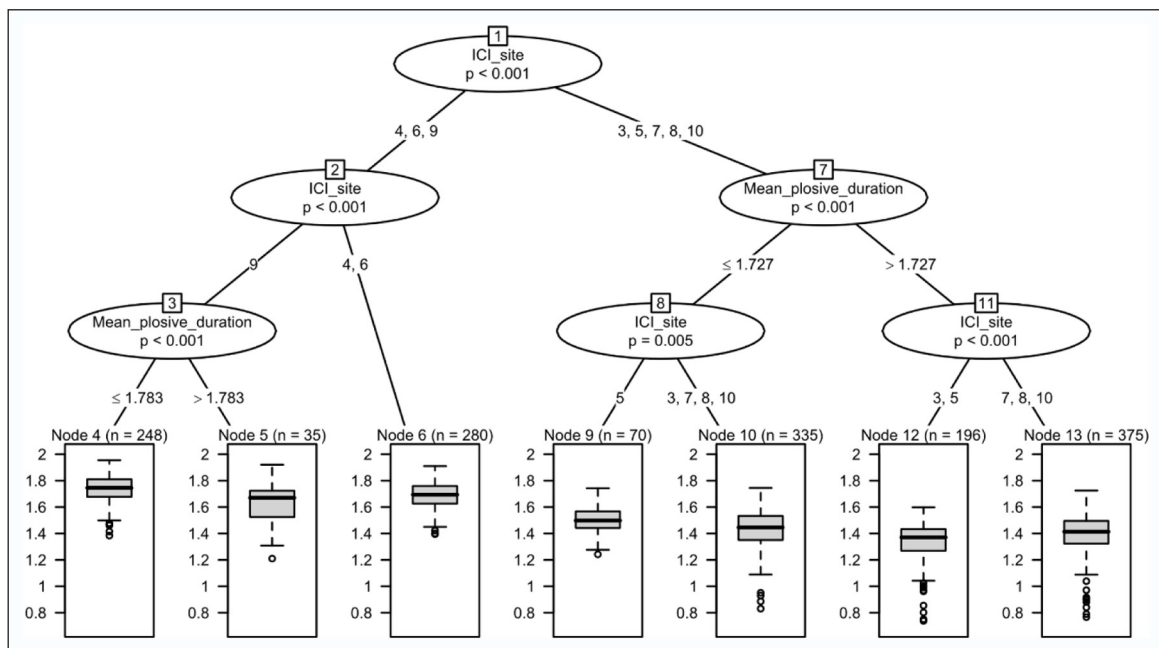


Figure 10: Conditional inference regression tree for ICI Duration (log ms) with ICI Site and Mean Plosive Duration (log ms) as conditioning variables; the box plots at the terminal nodes show the distributions of ICI Duration (log ms) in the relevant subsets of data.

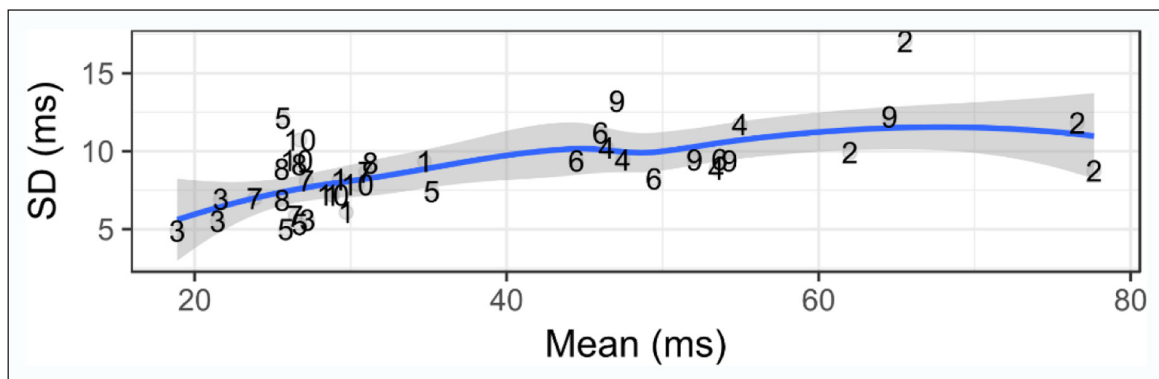


Figure 11: Scatter plot for ICI Duration (ms) means and standard deviations, by ICI Site (labels overlaying the data points) and Speaker (not labelled), with LOESS fit line and 95% confidence interval.

interval 0.40–0.79). The correlations across sites 2, 4, 6, and 9 alone are very similar to those across sites 1, 3, 5, 7, 8, and 10 alone ($r = 0.24$, 95% CI -0.29 – 0.66 and $r = 0.33$, 95% CI -0.08 – 0.65 respectively). Given that the overall correlation is not perfect, and given the clear separation of the two groups of sites in terms of duration means, sites 2, 4, 6, and 9 are characterized by comparatively low relative standard deviations, or ‘coefficients of variance’ (see Shaw et al., 2009; Shaw et al., 2011)—calculated by dividing the standard deviation for each site by the site’s mean duration. Across speakers, relative standard deviations for sites 2, 4, 6, and 9 are in the range 0.20–0.23; for sites 1, 3, 5, 7, 8, and 10 they are in the range 0.26–0.33. This may be taken to suggest that the former sites are characterized by relative temporal stability; however, we believe that not enough is known about the relationship between means and relative standard deviations in speech production to draw firm conclusions from this pattern.

3.5. Voice spreading

3.5.1. Descriptives

To investigate whether there is any evidence of ICIs at predicted epenthesis sites and at other sites interacting differently with [VOICE] spreading patterns, we first inspected the distributions of Relative Plosive Voicing Fraction by the plosives’ sequential positions (1, 2, 3, or 4, across sequences of two to four plosives). Recall that a Relative Plosive Voicing Fraction value of zero means the plosive’s voicing fraction is consistent with its underlying status; a value of -1 means an underlyingly voiced plosive surfaces with full voicelessness (i.e., is fully devoiced); and a value of 1 means an underlyingly voiceless plosive surfaces with full voicing (i.e., is fully voiced). **Figure 12** shows that plosives in first position mostly surface according to their underlying voicing specification (centre peak) or undergo voicing (right peak). Values below -0.5 , which reflect substantial devoicing of underlyingly voiced plosives, are very rare, and consistency across speakers

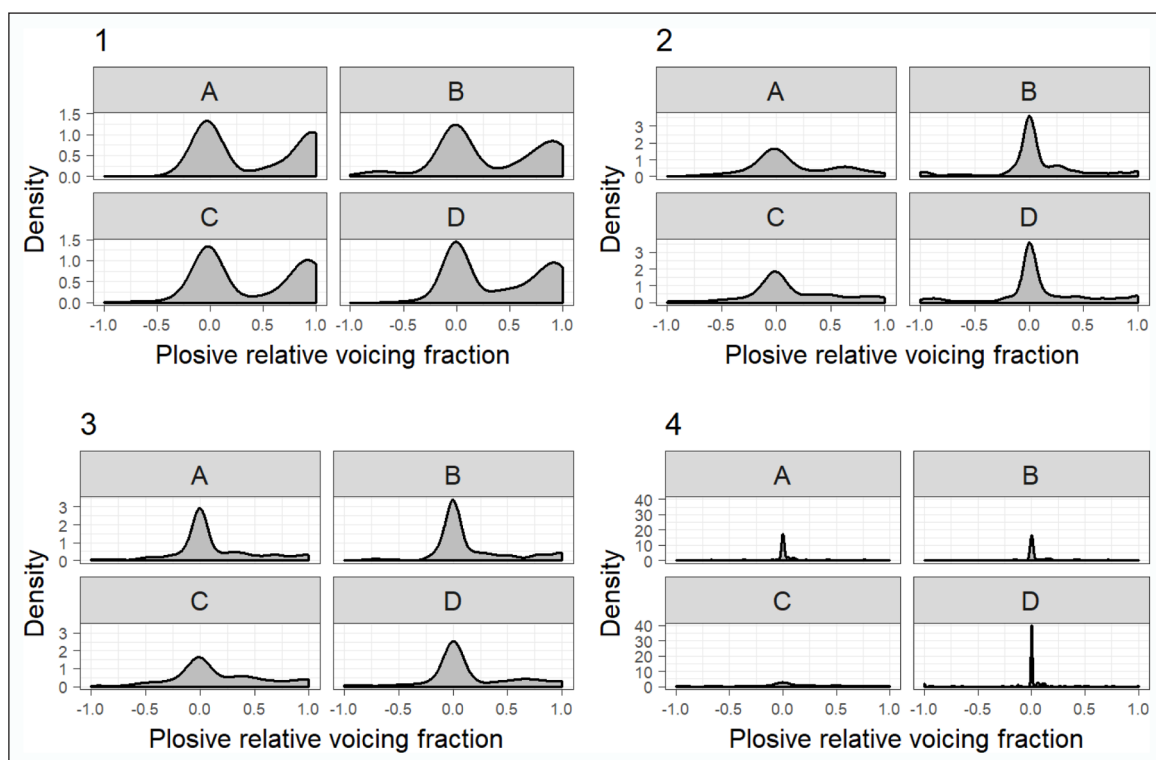


Figure 12: Kernel distribution plots for Plosive Relative Voicing Fraction by sequence position (1, 2, 3, 4), split by Speaker.

is high. Plosives in second, third, and fourth positions mostly surface according to their underlying specification. Surface voicing is less common here, but observed in sizeable subsets of the data, with some inter-speaker variation. Surface devoicing is again very rare. In sum, we find clear evidence for voicing of underlyingly voiceless plosives in TLA plosive sequences, and little evidence for devoicing of underlyingly voiced plosives occurring with any regularity.

3.5.2. /t/ and /k/ voicing by sequential position

Given the observations above, we focused our attention in modelling Plosive Relative Voicing Fraction on the underlyingly voiceless plosives in our data: /t/ and /k/. We proceeded by sequential position (1, 2, 3, 4). For each plosive, we fitted a linear mixed effects model containing random intercepts for Speaker and Repetition and fixed effects for the Plosive Voicing Fraction values of immediately adjacent plosives, ICI Site levels of immediately adjacent ICIs, and the identity of the plosive (/t/ or /k/). We removed any non-significant predictors, and then assessed, through inspection of alternative models' AIC values, whether replacing fixed effects by interactions improved model fit, and whether adding random slopes for Speaker and Repetition did. The latter was not the case for any of our models, and the random effect for Repetition did not contribute significantly to model fit in any. We also assessed whether replacing Plosive Voicing Fraction values of immediately adjacent plosives by corresponding Plosive Underlying Voicing improved model fit; this was to allow for the possibility that underlyingly voiced and voiceless plosives condition adjacent plosives differently, even if their own surface voicing is identical. In fact, this replacement reduced fit in all of our models. This is consistent with the prevalence of surface voicing throughout our plosive sequences and makes the identification of possible spreading 'triggers' rather hard. We return to this issue below; first we report the final models for /t/ and /k/ voicing fractions and present corresponding conditional inference regression trees for visualization. We do not report modelling for /t/ and /k/ in sequential position 4 (-CC#CC-) because the model for position 3 (-CC#CC-) is clear on the relationship between the third and fourth hold phases.

We modelled the voicing of /t/ and /k/ in position 1 across two, three, and four-plosive sequences (N = 545). A pertinent distinction is between two and four-plosive sequences on the one hand and three-plosive sequences on the other: In the latter (-VC₄#CCV-, -VC₃#CV-), the plosive hold phase is separated from the following plosive by a predicted epenthesis site; in the former (-VC₃#CV-, -VC₃#CCV-) it is not. An associated interaction between ICI Site and the voicing of the following plosive might provide evidence for predicted epenthesis sites and other sites influencing the extent of [VOICE] spreading differently. In fact, the best fit linear mixed effects model has fixed effects only: for the following ICI Site ($F(3,537) = 34.43, p < 0.001$) and for Plosive Voicing Fraction for the following hold phrase ($F(1,538) = 25.06, p < 0.001$). Inspection of level comparison coefficients suggests that in relation to ICI Site, the main explanatory factor is whether the plosive is part of a coda cluster (-VC₃#CV-, -VC₃#CCV-), or followed by a word boundary (-VC₃#CV-, -VC₄#CCV-). This grouping of ICI sites is reflected in a conditional inference regression tree model, as seen in **Figure 13**. (We residualized Plosive Relative Voicing Fraction for Speaker prior to tree modelling, as above.) The tree model also confirms that in both contexts there is the same positive relationship between the voicing fractions of the first and second hold phases.

For the purpose of modelling the voicing of /t/ and /k/ in position 2, we restricted our attention to three and four-plosive sequences (N = 420). For two-plosive sequences (-C₂#C-), the relationship between the two plosives is clear from the preceding analysis,

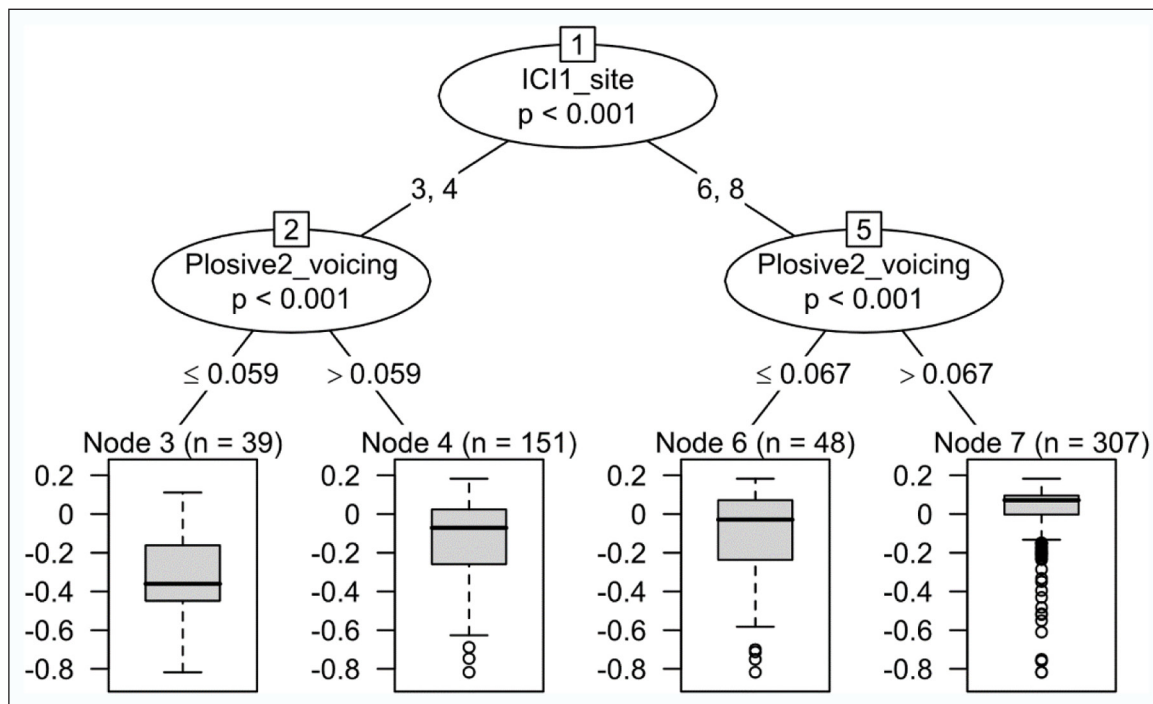


Figure 13: Conditional inference regression tree for Plosive Voicing Fraction for /t/ and /k/ in sequential position 1 (-C③#C-, -C④#CC-, -C⑥C#C-, -C③C#CC-), with the following ICI Site (here called ICI1 Site) and the voicing fraction for the following hold phase (Plosive2 Voicing) as conditioning variables; the box plots at the terminal nodes show the distributions of (residualized) Plosive Voicing Fraction in the relevant subsets of data.

and there is no third plosive to include in the model. The preceding and following ICI sites are logically related, such that modelling requires only one associated variable; we will refer to this as ICI Sites (levels ‘4_5’ for -C④#C⑤C-, ‘6_7’ for -C⑥C⑦#C-, and ‘8_9’ for -C③C⑨#CC-). In relation to the preceding ICI, sites 4 and 6 are predicted epenthesis sites; site 8 is not. In relation to the following ICI, site 9 is a predicted epenthesis site; sites 5 and 7 are not. We were again particularly interested in whether these groupings had any relevance for the relationships between the three successive plosives in hold phase voicing.

Again, however, the best fit linear mixed effects model has fixed effects only, for ICI Sites ($F(2,412) = 42.96, p < 0.001$), Plosive Voicing Fraction for the preceding hold phase ($F(1,398) = 37.84, p < 0.001$), and Plosive Voicing Fraction for the following hold phase ($F(1,414) = 8.72, p = 0.003$). Inspection of level comparison coefficients suggests that the grouping by ICI Sites separates -C④#C⑤C- and -C⑥C⑦#C- on the one hand from -C③C⑨#CC- on the other. In the former, the hold phase is preceded by a predicted ICI site, while in the latter, it is followed by one. In the latter context, hold phase voicing is less extensive. The same pattern is visible in a conditional inference regression tree model, as shown in **Figure 14**. The tree model further suggests that the effect of the voicing fraction of the following hold phase is only observed when that hold phase is *not* preceded by a predicted epenthesis site. If we interpret the positive correlation between the Plosive Voicing Fraction values for the second and third hold phases in terms of (leftward or rightward) spreading, this pattern might provide evidence for the predicted epenthesis site 9 (-CC⑨#CC-) limiting the extent of this spreading. We should note that site 9 is a word boundary site, and we have seen evidence for word boundaries constraining hold phase voicing relations in the analysis for sequential position 1 above. We will return to this below.

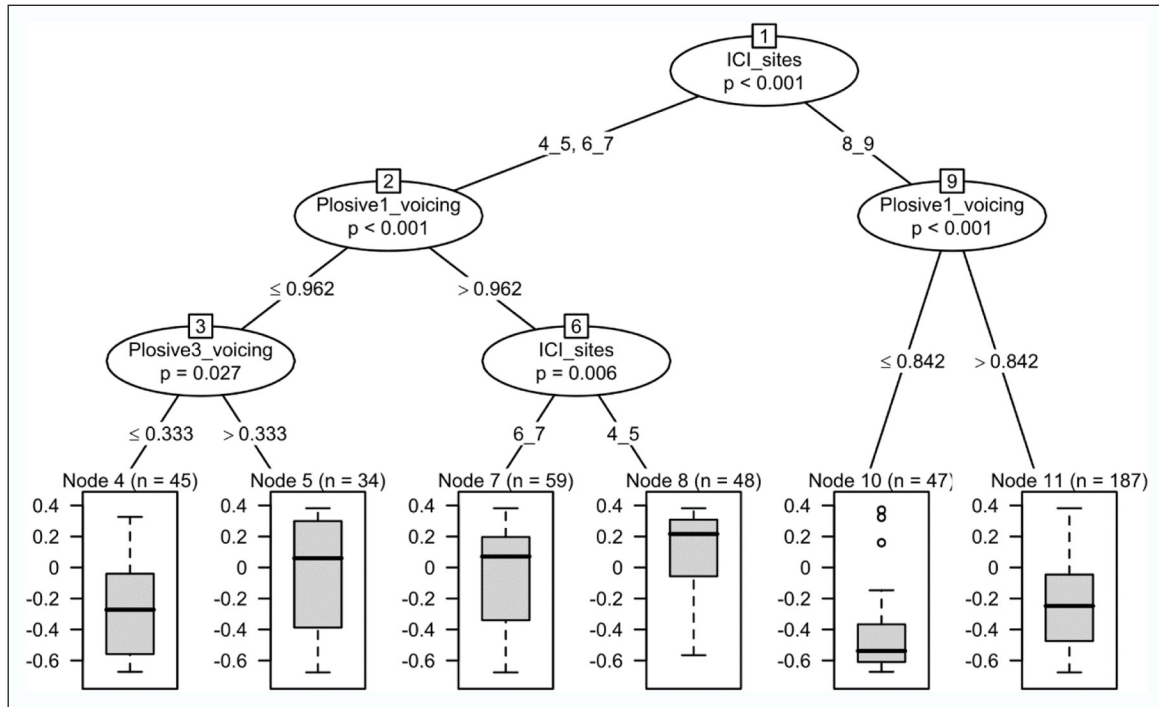


Figure 14: Conditional inference regression tree for Plosive Voicing Fraction for /t/ and /k/ in sequential position 2, with the preceding and following ICI Sites ('4_5' for -C $\text{\textcircled{4}}$ #C $\text{\textcircled{5}}$ -, '6_7' for -C $\text{\textcircled{6}}$ #C $\text{\textcircled{7}}$ -, '8_9' for -C $\text{\textcircled{8}}$ #C $\text{\textcircled{9}}$ CC-), the voicing fraction for the preceding hold phase (Plosive1 Voicing), and the voicing fraction for the following hold phase (Plosive3 Voicing) as conditioning variables; the box plots at the terminal nodes show the distributions of (residualized) Plosive Voicing Fraction in the relevant subsets of data.

For the purpose of modelling the voicing of /t/ and /k/ in third position, we restricted our attention to four-plosive sequences only (N = 248). Here (-CC $\text{\textcircled{4}}$ #C $\text{\textcircled{5}}$ C-), the hold phase is preceded by a predicted epenthesis site and word boundary; the following ICI site is not a predicted epenthesis site. The best fit model has fixed effects for both Plosive Voicing Fraction for the preceding hold phase ($F(1,245) = 7.69, p = 0.005$) and Plosive Voicing Fraction for the following hold phase ($F(1,245) = 40.49, p < 0.001$). The latter effect is the stronger of the two, consistent with the observation above that the correlation of voicing fraction values across ICI site 9 is relatively weak.

3.5.3. ICI and hold phase voicing

We now return to our earlier observation that while ICI sites 4, 6, and 9 are not distinguished by the occurrence of fully voiced ICIs between fully voiceless hold phases, they are associated with more substantial voicing in contexts with relatively little voicing. We illustrated this with the density scatters in **Figure 7**, which show a relative scarcity of data points in the bottom left corners for sites 4, 6, and 9. Another look suggests that more generally, low values for Mean Plosive Voicing Fraction are comparatively rare for these sites. This is primarily due to hold phases *following* sites 4, 6, and 9 having relatively high voicing fractions, as shown in **Figure 15**. (The effects of sites 4 and 6 are also reflected in the split for ICI Sites—'4_5' and '6_7' versus '8_9'—in the tree model in **Figure 14**.)

As we pointed out above, while we would not expect epenthetic vowels to block [+VOICE] spreading, it is possible for robustly voiced ICIs to give rise to voicing tails into subsequent underlyingly voiceless hold phases, thereby *contributing* to the spread of surface voicing. To check whether rightward spread of voicing from voiced ICIs might provide a reasonable account for the data pattern described here, we quantified rough hold phase

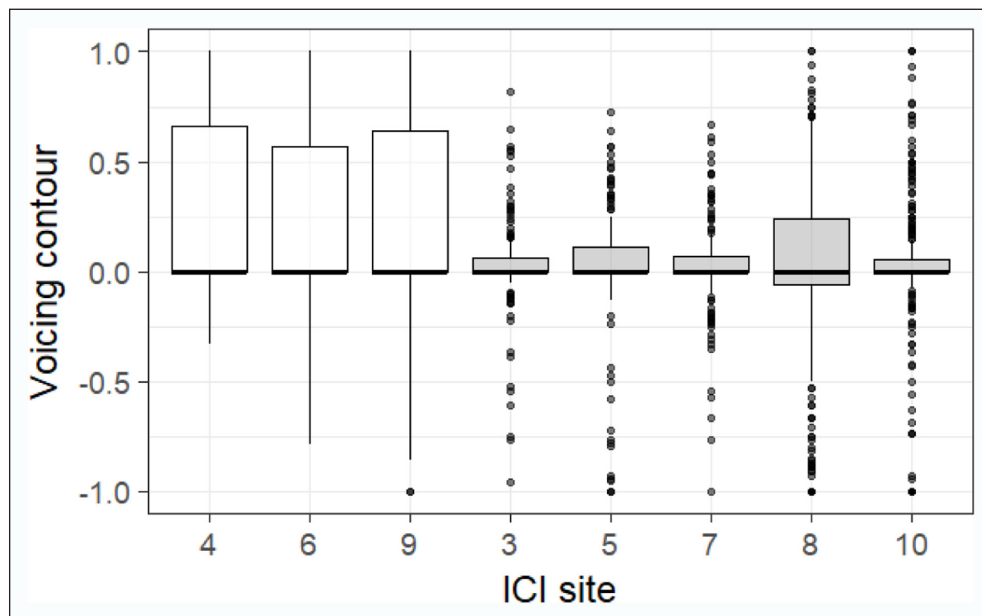


Figure 15: Box plots for Plosive2 Contour (positive values reflect derived voicing, negative values derived voicelessness; see text for explanation), split by the preceding ICI Site; boxes for the three predicted epenthesis sites (4, 6, 9) are in white on the left.

‘voicing contours’ for partially voiced instances of /t/ and /k/ (that is, instances with voicing fractions between 0.1 and 0.9, in line with Davidson, 2016), by subtracting the voicing fraction for the second half from that for the first half. On the resulting scale, a value of 1 means all voicing is located in the first half of the hold phase; zero means the (partial) voicing is evenly dispersed across the hold phase; and -1 means all voicing is in the second half. We then modelled the resulting variables, which we will call Plosive2 Contour (for /t/ and /k/ in sequential position 2) and Plosive3 Contour (for /t/ and /k/ in sequential position 3) to assess the relevance of the preceding ICI site.

Mixed effects modelling of Plosive2 Contour with Speaker as a random intercept and preceding ICI Site as a fixed factor reveals a significant effect of the latter ($F(2,277) = 7.31, p < 0.001$) due to a grouping of sites 4 and 6 on the one hand and 8 on the other. The same modelling of Plosive3 Contour also reveals a significant effect of preceding ICI Type ($F(2,164) = 7.08, p = 0.001$), due to a grouping of sites 5 and 7 on the one hand and 9 on the other. Conditional inference regression tree models show the same effects, as seen in **Figure 16**. (We residualized Plosive2 Contour and Plosive3 Contour for Speaker prior to tree modelling, as above.) Plosive2 Contour values are higher following ICI sites 4 and 6 (-C⁴#C⁵C-, -C⁶C⁷#C-) than following ICI site 8 (-C⁸C⁹#CC-), and Plosive3 contour values are higher following ICI site 9 (-CC⁹#C¹⁰C-) than following ICI sites 5 and 7 (-C#C⁵C-, -CC⁷#C-). This means /t/ and /k/ hold phases following the ‘predicted epenthesis’ sites 4, 6, and 9 are characterized by a relative preponderance of early voicing, consistent with the source of voicing being to the left rather than the right of the hold phase.

3.5.4. Summary of voice spreading analysis

While our aim in this study was not to provide an in-depth analysis of [VOICE] spreading in TLA, we can make an attempt to draw our findings together into a coherent, if tentative, account. Our analysis of /t/ and /k/ voicing in sequential position 1 (see **Figure 13**) suggested that voicing relations between first and second hold phases are constrained by intervening word boundaries. The finding of ICI site 9 (-CC⁹#CC-) constraining the voicing relations between second and third hold phases (see **Figure 14**) seems consistent with a

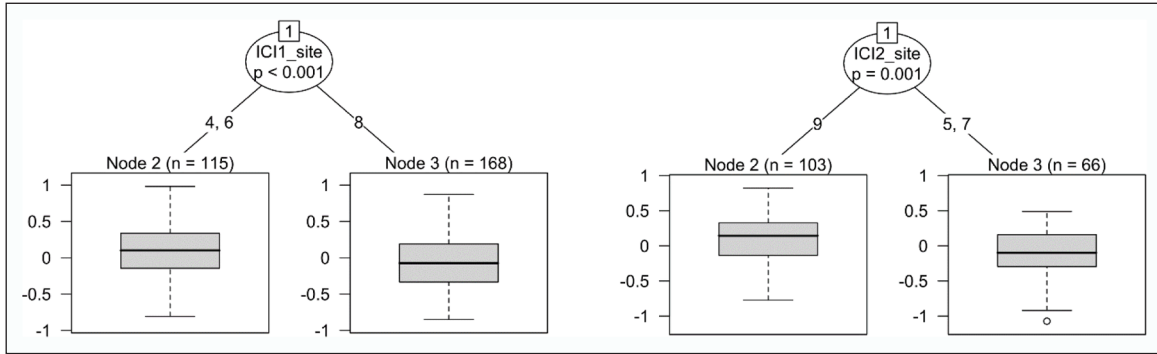


Figure 16: Conditional inference regression trees for (left) Plosive2 Contour (/t/ and /k/ in sequential position 2: -C⁴#C⁵C-, -C⁶C⁷#C-, -C⁸C⁹#CC-) and (right) Plosive3 Contour (/t/ and /k/ in sequential position 3: -C#C⁶C-, -CC⁷#C-, -CC⁹#C¹⁰C-), with the preceding ICI Site (here called ICI1 Site and ICI2 Site respectively) as conditioning variables; the box plots at the terminal nodes show the distributions of (residualized) Plosive2 Contour and Plosive3 Contour in the relevant subsets of data.

word boundary effect. This reasoning would appear to predict that ICIs at site 7 (-CC⁷#C-) should also constrain the voicing relations between second and third hold phases; however, as seen in **Figure 14**, these hold phases (-CC⁷#C-, labelled ‘6_7’) do not pattern with those preceding site 9 (-CC⁹#CC- labelled ‘8_9’) in relation to the voicing of the following hold phase. Having established that hold phases preceding site 7 (that is, following site 6) are characterized by a relative preponderance of early voicing, we can suggest that for these hold phases, a constraining effect of the following word boundary is masked by voicing coming in from ICI site 6 on the left. On this tentative interpretation of the data patterns described above, we find some evidence for word boundaries limiting the extent of [+VOICE] spreading (cf. Cho, 2001; Mohanan, 1993), and we find some evidence for the predicted epenthesis sites 4, 6, and 9 giving rise to voicing tails in following hold phases.

4. Discussion

This study was motivated by the observation that TLA is characterized by widespread, partly optional vowel epenthesis in consonant strings (Watson, 2007); we aimed to establish whether acoustic and distributional analyses of ICIs in plosive sequences provide insight into the nature of this ‘variable epenthesis.’ Here we summarize our findings and spell out some of their implications. For reference, we repeat the ICI sites we considered in (2). On the basis of Watson’s description of TLA, one would expect ICIs at sites 4, 6, and 9 to stand out from those at other sites in being associated with (near-)obligatory epenthesis, although we argued above that in principle, an account in terms of vocoid intrusion alone could be descriptively adequate.

- (2) a. ##C¹C-
- b. -C²C##
- c. -C³#C-
- d. -C⁴#C⁵C-
- e. -C⁶C⁷#C-
- f. -C⁸C⁹#C¹⁰C-

4.1. Evidence for phonological vowel insertion

We start with the question of whether we find compelling evidence in favour of a phonological epenthesis rule. We suggested above that we can count several data patterns as providing such evidence; we will take them in turn. First, we pointed out that an

observation of consistently voiced inter-consonantal intervals with clear formant structure between voiceless consonants is difficult to account for in terms of vocoid intrusion. Our analysis has suggested that our data set contains no fully voiced ICIs between two fully voiceless hold phases, and very few fully voiced ICIs between hold phases whose voicing fractions average below 0.5. There is a tendency for substantial ICI voicing, at voicing fractions around 0.8, in predicted epenthesis sites (4, 6, and 9) when surrounding /t/ and /k/ hold phase voicing fractions are considerably lower, around or below 0.5. An account including vowel epenthesis would clearly explain this. It might have to accommodate partial devoicing of epenthetic vowels, however, as low ICI voicing fractions do occur at these sites—although much less frequently than at sites 3, 5, 7, 8, and 10.

Second, we pointed out that an observation of bimodality in the distributions of inter-consonantal interval durations is difficult to account for in terms of vocoid intrusion, unless it can be shown to fall out of the interplay of general, independently motivated factors influencing inter-consonantal timing. We found significant bimodality in ICI durations for one speaker, and similar distributions for all, with a main peak centred around 25 ms and a ‘right tail’ centred around or above 50 ms. We saw no sign of similar non-normality when we considered durations by ICI site: Its source was significantly higher ICI durations in the three predicted ICI sites plus ICI site 2. The latter pattern was observed for all speakers.

Note that we have been careful to emphasize that bimodality points towards an account including epenthesis if it cannot be attributed to the interplay of independently motivated factors influencing inter-consonantal timing. We believe that it is hard to make this attribution in our data set. Sites 2, 4, 6, and 9 include two word-boundary sites (4 and 9), at which gestural coordination is expected to be relatively loose; however, it also includes two coda sites (2 and 6), at which gestural coordination is expected to be tighter, and this is not reflected in the ICI durations. Moreover, the third word boundary site in our data, site 3, is *not* associated with particularly long ICIs. Crucially, the observation that ICIs at sites 4 and 6 are not significantly different in their phonetic characteristics fits well with Watson’s characterization of TLA as a ‘VC’ dialect of Arabic, and makes an account in terms of general timing constraints problematic: These sites have in common that they follow the first plosive of a three-plosive sequence, but are otherwise phonologically distinct—and the fact that other dialects of Arabic are associated with different distributions of phonetic ICI types across these sites entails a degree of arbitrariness that seems hard to accommodate in an account that appeals to general timing constraints only.

In fact, our cluster analysis suggested that ICIs at the three predicted epenthesis sites and site 2 have distinct phonetic profiles across the parameters of voicing, duration, and amplitude, compared with ICIs elsewhere: The former are associated with high voicing fraction values, high duration values, and mid or high amplitude values depending on the speaker, while the latter are associated with variable voicing, low duration values, and low-to-mid amplitude values. These profiles seem a good fit for an epenthetic vowel versus intrusive vocoid distinction as outlined by Hall (2006): Epenthetic vowels are more commonly voiced, longer, and produced with more intensity compared with apparent vowel portions resulting from variable open transition. We interpret the distributions of our acoustic parameters and the outcome of our modelling of cluster membership as suggesting that if phonological vowel insertion is at play in sites 2, 4, 6, and 9, it is the norm: In particular, ICIs belonging to the cluster with variable voicing, low duration, and low amplitude are a very small minority in these sites. Moreover, if vowel insertion is at play at all in other sites, it is not widespread: ICIs belonging to the clusters with high voicing, high duration, and mid-to-high amplitude account 20% to 30% of instances in sites 1, 3, 5, 7, 8, and 10, depending on the speaker. In other words, variable open transition with regular vocoid intrusion seems the norm in the latter sites.

Third, we pointed out that an observation of consistently lower rates of phonological process application—in our case we examined [VOICE] spreading—across ICIs in particular contexts is difficult to account for without appealing to the ‘blocking’ potential of epenthetic vowels. Our analysis of plosive hold phase voicing suggested that [-VOICE] spreading is very rare in TLA plosive sequences, so we could not assess the blocking potential of the predicted epenthesis sites. We did find evidence for extensive surface voicing of /t/ and /k/. We found that plosive hold phases are more extensively voiced following the three predicted epenthesis sites 4, 6, and 9 than elsewhere, and the voicing tends to be early rather than late. We have interpreted this pattern in terms of the ICIs giving rise to subsequent voicing tails. If this interpretation is reasonable, we can tentatively conclude that ICIs at predicted epenthesis sites spread voicing themselves. This would seem hard to account for without an appeal to phonological vowel insertion.

4.2. Other inter-consonantal timing patterns

While the findings summarized so far point in the direction of an account of TLA variable epenthesis that includes epenthesis proper, we did *not* find clear supporting evidence in several areas. First, we pointed out that a lack of correlation between ICI duration means and standard deviations can be taken as evidence in favour of an account including vowel insertion. We found that in our data set, ICI duration means and standard deviations are strongly correlated across ICI sites, which would be expected if only vocoid intrusion was at play. Whether the correlation provides evidence *against* an account incorporating epenthesis is another matter, and we believe too little is known about this relationship for lexical segments to assume that it does.

Second, we pointed out that a lack of negative correlation between the durations of ICIs and surrounding hold phases can be taken as evidence in favour of an account including vowel insertion: In previous studies, epenthetic vowels have been found to add duration to consonant clusters, while intrusive vocoids do not. We found a lack of negative correlation between the durations of ICIs and surrounding hold phases for two out of the three predicted epenthesis sites—sites 4 and 6—but found the remaining site—site 9—indistinguishable from sites in which epenthesis is not predicted to be widespread.

Third, we argued that place order effects might be expected to be more robust in those sites in which ICIs presumably consist largely of open transitions—that is, sites other than our predicted epenthesis sites 4, 6, and 9. We found no evidence for this: Instead, place order effects are observed across all sites except those with the shortest and longest ICI durations; crucially, they appear equal in extent across sites 4, 6, and 9 on the one hand and sites 1, 5, 7, 8, and 10 on the other. Their non-appearance in site 2 can be explained in terms of the ‘relativized place order hypothesis’ of Gafos et al. (2010): The longer ICIs are, the less likely it is that they are subject to place order effects. Gafos et al. (2010) do not refer to very short ICIs; our findings suggest there may be a lower as well as an upper limit to the range of ICI durations in which place order effects can be observed to an extent that reaches statistical significance.

The upshot of these findings is that if we accept that epenthesis is at play in sites 2, 4, 6, and 9, we must also accept that the consonant sequences in which this epenthesis takes place have temporal characteristics that make them hard to distinguish from consonant sequences in which epenthesis is most likely *not* at play. We can rule out the possibility that these observed characteristics are due to the small subsets of ICIs at these sites with variable voicing, low duration, and low amplitude: Modelling ICI Duration (log ms) with members of the associated cluster (cluster 1 in the analysis above) removed reveals the same effects of Mean Hold Phase Duration and Place Order as those summarized above

(modelling details not shown). The resulting picture seems at odds with Hall's (2006) account of vowel intrusion, which implies that intrusive vocoids and epenthetic vowels are distinct objects, and that epenthetic vowels have surface representations identical to those of lexical vowels. It has been beyond the scope of the current study to investigate lexical vowels in TLA; it would be interesting to compare the characteristics of the ICIs at sites 2, 4, 6, and 9 to those of lexical vowels embedded in similar plosive sequences. We can note that Hall (2013) shows that for some speakers of Lebanese Arabic, epenthetic vowels are qualitatively different from lexical vowels, and Blevins and Pawley (2010) question the validity of Hall's binary distinction between vocoid intrusion and vowel epenthesis, based on an analysis of 'predictable vowels' in Kalam.

We should also note that we cannot offer a clear explanation of why ICIs at site 2, in utterance-final codas, pattern with those at the 'predicted epenthesis' sites 4, 6, and 9. The answer may lie in constraints on the prosodic structure of content words, for example that they must form at least one bimoraic foot, the second mora being supplied utterance-finally by epenthesis in the case of final consonants being prosodically weightless (Hayes, 1989). We consider the development of such an account beyond the scope of this paper.

4.3. Methodological contribution

Finally, our analysis has demonstrated that vowel epenthesis and intrusion can be fruitfully investigated without recourse to native speaker intuitions as to the status of particular inter-consonantal intervals. We believe this is a notable strength of our work *per se*: It means our account is to be assessed in terms of the rigour of our acoustic and phonological analyses, not to the presumed reliability of our speakers' judgements. It also means our analysis can be replicated with relatively few unknown 'degrees of freedom (Roettger, 2018).' Similarly, we would argue that studies for which it is important to distinguish between consonant releases on the one hand and epenthetic or intrusive vowels on the other hand should in the first instance assess whether this distinction emerges from relevant acoustic measurements. If it does, as in the case of Wilson and Davidson (2015), the acoustic analysis most likely obviates the need for expert coding—at least in the context of a production study. If it does not, expert coding is not likely to produce informative results. We therefore hope to have made contributions to both the study of vowel epenthesis and the study of Arabic consonant sequences, as well as a valid methodological contribution to the study of inter-consonantal coordination more generally.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** This document lists the 111 lexical items produced by our speakers, in phonemic transcription, with gloss and associated ICI site(s). DOI: <https://doi.org/10.5334/labphon.122.s1>

Acknowledgements

We are grateful to Aimen Ghummed for his early contribution to this research, to Janet Watson for helpful discussion, to Chris Norton for technical assistance, and to our speakers for giving us their time. We are indebted to Lisa Davidson and two anonymous *Laboratory Phonology* reviewers for their detailed criticism of the first submission of this paper—and for allowing us to revise and resubmit. We are grateful to Cécile De Cat and Patrycja Strycharczuk for taking the time to read and comment on subsequent drafts.

Competing Interests

The authors have no competing interests to declare.

References

- Al-Deaibes, M. 2016. *The phonetics and phonology of assimilation and gemination in Rural Jordanian Arabic*. Unpublished Doctoral Dissertation, University of Manitoba.
- Baayen, R. H. 2013. Multivariate statistics. In: Podesva, R. J., & Sharma, D. (eds.), *Research methods in linguistics*, 337–372. Cambridge: Cambridge University Press.
- Barry, M., & Teifour, R. 1999. *Temporal patterns in Syrian Arabic voicing assimilation. Paper presented at the 14th International Congress of Phonetic Sciences*. San Fransisco.
- Bates, D., Maechler, M., Bolker, B. M., & Walker, S. C. 2015. Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Bayles, A., Kaplan, A., & Kaplan, A. 2016. Inter- and intra-speaker variation in French schwa. *Glossa-a Journal of General Linguistics*, 1(1). DOI: <https://doi.org/10.5334/gjgl.54>
- Bellik, J. 2018. An acoustic study of vowel intrusion in Turkish onset clusters. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 9(1), 1–23. DOI: <https://doi.org/10.5334/labphon.112>
- Blevins, J., & Pawley, A. 2010. Typological implications of Kalam predictable vowels. *Phonology*, 27, 1–44. DOI: <https://doi.org/10.1017/S0952675710000023>
- Boersma, P., & Weenink, D. 2016. Praat: Doing phonetics by computer (Version 6.0.21).
- Broselow, E. 1992. Parametric variation in Arabic dialect phonology. In: Broselow, E., Eid, M., & McCarthy, J. J. (eds.), *Perspectives on Arabic linguistics IV*, 7–45. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.85.04bro>
- Browman, C. P., & Goldstein, L. 1992. ‘Targetless’ schwa: An articulatory analysis. In: Docherty, G., & Ladd, D. R. (eds.), *Papers in Laboratory Phonology II: Gesture, segment, prosody*, 26–56. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511519918.003>
- Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2), 209–244. DOI: <https://doi.org/10.1006/jpho.1996.0012>
- Catford, J. C. 2001. *A practical introduction to phonetics* (2nd edition ed.). Oxford: Oxford University Press.
- Chitoran, I., Goldstein, L., & Byrd, D. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory Phonology*, 7(4–1), 419–447. DOI: <https://doi.org/10.1515/9783110197105.419>
- Cho, T. 1998. *Specification of intergestural timing and gestural overlap: EMA and EPG studies*. Unpublished MA thesis, University of California.
- Cho, T. 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica*, 58(3), 129–162. DOI: <https://doi.org/10.1159/000056196>
- Clements, G., & Sezer, E. 1982. Vowel and consonant disharmony in Turkish. In: Van der Hulst, H., & Smith, N. (eds.), *The structure of phonological representations*. Dordrecht: Foris.
- Coleman, J. 2001. The phonetics and phonology of Tashlhiyt Berber syllabic consonants. *Transactions of the Philological Society*, 99(1), 29–64. DOI: <https://doi.org/10.1111/1467-968X.00073>
- Davidson, L. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54, 35–50. DOI: <https://doi.org/10.1016/j.wocn.2015.09.003>
- Davidson, L. 2018. Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association*, 48(3), 331–356. DOI: <https://doi.org/10.1017/S0025100317000330>

- Davis, S. 1995. Emphasis spread in Arabic and grounded phonology. *Linguistic Inquiry*, 26(3), 465–498.
- Dell, F., & Elmedlaoui, M. 2002. *Syllables in Tashliyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer. DOI: <https://doi.org/10.1007/978-94-010-0279-0>
- Eager, C. D. 2015. Automated voicing analysis in Praat: Statistically equivalent to manual segmentation. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.
- Gafos, A. I. 2002. A grammar of gestural coordination. *Natural Language & Linguistic Theory*, 20(2), 269–337. DOI: <https://doi.org/10.1023/A:1014942312445>
- Gafos, A. I., Hoole, P., Roon, K., & Zeroual, C. 2010. Variation in overlap and phonological grammar in Moroccan Arabic clusters. In: Fougeron, C., Kühnert, B., d'Imperio, M., & Vallé, N. (eds.), *Laboratory Phonology 10*, 657–698. Berlin: Mouton de Gruyter.
- Ghummed, A. 2015. *An acoustic and articulatory analysis of consonant sequences across word boundaries in Tripolitanian Libyan Arabic*. Unpublished doctoral dissertation, University of Leeds.
- Hall, N. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology*, 23(3), 387–429. DOI: <https://doi.org/10.1017/S0952675706000996>
- Hall, N. 2011. Vowel epenthesis. In: Van Oostendorp, M., Ewen, C. J., Hume, E., & Rice, K. (eds.), *The Blackwell Companion to Phonology*, 1576–1596. Malden, MA: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444335262.wbctp0067>
- Hall, N. 2013. Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic. *Journal of Phonetics*, 41(2), 133–143. DOI: <https://doi.org/10.1016/j.wocn.2012.12.001>
- Hayes, B. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20, 253–306.
- Heath, J. 1987. *Ablaut and ambiguity: Phonology of a Moroccan Arabic dialect*. Albany: State University of New York Press.
- Hothorn, T., & Zeileis, A. 2015. Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16, 3905–3909.
- Jansen, W. 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian and Dutch*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen.
- Kabrah, R. 2011. Regressive voicing assimilation in Cairene Arabic. In: Broselow, E., & Ouali, H. (eds.), *Perspectives on Arabic Linguistics XXII–XXIII*, 21–33. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.317.02kab>
- Kaufman, L., & Rousseeuw, P. J. 1990. *Finding groups in data: An introduction to cluster analysis*. New York: Wiley. DOI: <https://doi.org/10.1002/9780470316801>
- Kiparsky, P. 2003. Syllables and moras in Arabic. In: Féry, C., & Van der Vijver, R. (eds.), *The Syllable in Optimality Theory*, 147–182. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511497926.007>
- Liaw, A., & Wiener, M. 2002. Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Maechler, M. 2015. Diptest: Hartigan's dip test statistic for unimodality—corrected (Version 0.75-7).
- Maechler, M., Rousseeuw, P. J., Struyf, A., Hubert, M., & Hornik, K. 2016. Cluster: Cluster analysis basics and extensions (Version 2.0.5).
- McPherson, L., & Hayes, B. 2016. Relating application frequency to morphological structure: The case of Tommo So vowel harmony. *Phonology*, 33(1), 125–167. DOI: <https://doi.org/10.1017/S0952675716000051>


- Mohanan, K. 1993. Fields of attraction. In: Goldsmith, J. (ed.), *The last phonological rule: Reflections on constraints and derivations*, 61–116. Chicago: Chicago University Press.
- Odden, D. 2013. *Introducing phonology* (2nd edition ed.). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139381727>
- Owens, J. 2006. *A linguistic history of Arabic*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199290826.001.0001>
- Plug, L., & Carter, P. 2013. Prosodic marking, pitch and intensity in spontaneous lexical self-repair in Dutch. *Phonetica*, 70(3), 155–181. DOI: <https://doi.org/10.1159/000355512>
- Plug, L., & Carter, P. 2014. Timing and tempo in spontaneous phonological error repair. *Journal of Phonetics*, 45, 52–63. DOI: <https://doi.org/10.1016/j.wocn.2014.03.007>
- R Core Team. 2016. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Ridouane, R. 2008. Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology*, 25(2), 321–359. DOI: <https://doi.org/10.1017/S0952675708001498>
- Ridouane, R., & Fougeron, C. 2011. Schwa elements in Tashlhiyt word-initial clusters. *Laboratory Phonology*, 2(2), 275–300. DOI: <https://doi.org/10.1515/labphon.2011.010>
- Roettger, T. B. 2019. Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10(1), 1. DOI: <http://doi.org/10.5334/labphon.147>
- Schoner, G. 2002. Timing, clocks, and dynamical systems. *Brain and Cognition*, 48(1), 31–51. DOI: <https://doi.org/10.1006/brcg.2001.1302>
- Shaw, J. A., & Davidson, L. 2011. Perceptual similarity in input-output mappings: A computational/experimental study of non-native speech production. *Lingua*, 121, 1344–1358. DOI: <https://doi.org/10.1016/j.lingua.2011.03.003>
- Shaw, J. A., Gafos, A. I., Hoole, P., & Zeroual, C. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology*, 26(1), 187–215. DOI: <https://doi.org/10.1017/S0952675709001754>
- Shaw, J. A., Gafos, A. I., Hoole, P., & Zeroual, C. 2011. Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology*, 28(3), 455–490. DOI: <https://doi.org/10.1017/S0952675711000224>
- Shaw, J. A., & Kawahara, S. 2018. The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics*, 66, 100–119. DOI: <https://doi.org/10.1016/j.wocn.2017.09.007>
- Shitaw, A. E. 2014. *An instrumental phonetic investigation of timing relations in two-stop consonant clusters in Tripolitanian Libyan Arabic*. Unpublished doctoral dissertation, University of Leeds.
- Strobl, C., Malley, J., & Tutz, G. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348. DOI: <https://doi.org/10.1037/a0016973>
- Strycharczuk, P., Van 't Veer, M., Bruil, M., & Linke, K. 2014. Phonetic evidence on phonology-morphosyntax interactions: Sibilant voicing in Quito Spanish. *Journal of Linguistics*, 50(2), 403–452. DOI: <https://doi.org/10.1017/S0022226713000157>
- Surprenant, A. M., & Goldstein, L. 1998. The perception of speech gestures. *Journal of the Acoustical Society of America*, 104(1), 518–529. DOI: <https://doi.org/10.1121/1.423253>
- Tagliamonte, S. A., & Baayen, R. H. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. DOI: <https://doi.org/10.1017/S0954394512000129>

- Teifour, R. 1997. *Some phonetic and phonological aspects of connected speech processes in Syrian Arabic*. Unpublished doctoral dissertation, University of Manchester.
- Tomaschek, F., Hendrix, P., & Baayen, R. H. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71, 249–267. DOI: <https://doi.org/10.1016/j.wocn.2018.09.004>
- Torreira, F., & Ernestus, M. 2011. Vowel elision in casual French: The case of vowel /e/ in the word *c'était*. *Journal of Phonetics*, 39(1), 50–58. DOI: <https://doi.org/10.1016/j.wocn.2010.11.003>
- Vaux, B. 2003. Syllabification in Armenian, universal grammar, and the lexicon. *Linguistic Inquiry*, 34(1), 91–125. DOI: <https://doi.org/10.1162/002438903763255931>
- Watson, J. C. E. 2002. *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
- Watson, J. C. E. 2007. Syllabification patterns in Arabic dialects: Long segments and mora sharing. *Phonology*, 24(2), 335–356. DOI: <https://doi.org/10.1017/S0952675707001224>
- Wayland, R., & Jongman, A. 2003. Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, 31(2), 181–201. DOI: [https://doi.org/10.1016/S0095-4470\(02\)00086-4](https://doi.org/10.1016/S0095-4470(02)00086-4)
- Wilson, C., & Davidson, L. 2015. Acoustic characteristics of open transition in nonnative consonant cluster production. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Wilson, C., Davidson, L., & Martin, S. 2014. Effects of acoustic-phonetic detail on cross-language speech production. *Journal of Memory and Language*, 77, 1–24. DOI: <https://doi.org/10.1016/j.jml.2014.08.001>
- Zsiga, E. C. 1994. Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, 22(2), 121–140.

How to cite this article: Plug, L., Shitaw, A., and Heselwood, B. 2019 Inter-consonantal intervals in Tripolitanian Libyan Arabic: Accounting for variable epenthesis. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1):5, pp.1–33. DOI: <https://doi.org/10.5334/labphon.122>

Submitted: 13 October 2017 **Accepted:** 07 February 2019 **Published:** 05 March 2019

Copyright: © 2019 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 