Chapter 8

Intrusive and epenthetic vowels revisited

Nancy Hall
California State University Long Beach

The distinction drawn in Hall (2006) between epenthetic and intrusive vowels has been widely used in descriptive and theoretical work, and also challenged. This paper reviews some of the work that has engaged with the theory, in agreement or disagreement. This includes experimental approaches; proposed extensions to vowels inserted in other positions; and work arguing that vowels differ along a continuum rather than falling into these two classes. I argue that the dichotomy of presence/absence of a gesture is still a useful concept, while acknowledging that this is not the only important distinction among inserted vowels. An expanded typology recognizes subclasses among both intrusive and epenthetic vowels, as well as the possibility of alternation between epenthesis and intrusion.

1 Introduction

Vowel insertion is a cross-linguistically ubiquitous phenomenon – it is doubtful that any language does not display it in some way, at least in loanword adaptation or learner speech. But vowel insertion processes are diverse, in both their phonetic substance and their phonological patterning. Inserted vowels may be phonetically indistinguishable from lexical vowels, or similar but not identical to lexical vowels, or short, variable and indistinct. Some vowel insertion processes are optional, others obligatory. In terms of phonological behavior, some inserted vowels are clearly incorporated into the syllabic and metrical structure of the word, interacting with allophony, stress assignment, and intonation, while other inserted vowels are ignored by some or all aspects of the phonology. Some are fully accessible to the metalinguistic consciousness of speakers, as reflected

for example in spelling, while in other cases native speakers seem unaware that they produce a vowel, despite it being clearly audible to linguists and visible on spectrograms. Describing and explaining this diversity is one of the tasks of phonological theory.

In Hall (2006), I argued that one important distinction among vowel insertion processes is whether the perceived vowel corresponds to a distinct articulatory gesture. This analysis is framed in the theory of Articulatory Phonology (Browman & Goldstein 1992b), in which abstract articulatory gestures are units of phonological representation. In VOWEL EPENTHESIS, speakers insert into the phonological representation a new vowel gesture, which was absent underlyingly and/or historically. In VOWEL INTRUSION, no new gesture is inserted, but a low degree of overlap between adjacent consonant gestures produces an open transition that can sound like a vowel. The class of intrusive vowels includes the short and indistinct vowels often called "excrescent" (Levin 1987), but also includes some vowels that are relatively long and distinct in quality.

This proposal was partly spurred by the popularity of typological studies of insertion in Optimality Theory (e.g., Kitto & De Lacy 1999, Fleischhacker 2005, among many others). Cross-linguistic variation along particular dimensions, such as inserted vowel quality or location of insertion, was often cited as evidence for universal constraints. However, some typologies mixed together insertion processes that, on reading of the descriptive sources, seemed like qualitatively different phenomena. In the push to capture all of phonology with a single set of constraints, the distinction between more and less "phoneticky" processes was sometimes ignored. The epenthesis/intrusion distinction was a way to bring this back.

Since then, the intrusive/epenthetic distinction has since been widely used in the analysis of vowel insertion phenomena. Examples of languages claimed[1] to have intrusive vowels include Albanian (Canalis 2007), Alsea (Buckley 2007), Qaqet (Tabain & Hellwig 2022), Khmer (Butler 2015), Norwegian (Garmann et al. 2021), ancient Carian (Adiego 2019), Pnar (Ring & Gruber 2014), as well as a large number of examples in Easterday (2017)'s study of highly complex consonant clusters. Examples of languages claimed to have both epenthetic and intrusive vowels (with different characteristics) include Tripolitanian Libyan Arabic (Heselwood et al. 2015), Turkish (Bellik 2019, 2024 [this volume]), SENĆOŦEN (Leonard 2007), and Sephardic Hebrew (Pariente 2010). Sometimes intrusion vs

---

[1]These are offered as examples of how the theory has been applied; I do not necessarily agree with all of the classifications mentioned in this paragraph. Also, note that some of these sources use the term "excrescent vowel" rather than "intrusive vowel", but adopt an explicitly gestural definition of excrescence that is in line with what I call intrusion.

epenthesis is posited as a difference between closely related dialects, as in Cavirani (in preparation)'s study of Carrarese and Pontremolese Italian. The distinction has also been applied to vowel insertion in L2 learner speech (Nogita & Fan 2012) and aphasic speech (Buchwald et al. 2006).

However, the dichotomy has also been challenged as too rigid, too simplistic, or simply not a useful framework. Blevins & Pawley (2010) argue that "a simple two-way division between intrusive phonologically invisible vowels and epenthetic phonologically visible vowels is too restrictive", preferring to classify inserted vowels primarily by their historical pathways of development. Hammond et al. (2014) suggest there is a continuum of vowel types, rather than a categorical division.

In this paper I review some of the recent work that has engaged with the theory – whether in agreement or disagreement. I argue that the dichotomy of presence/absence of a gesture is still a useful concept, and can explain some of the variation among inserted vowels, while acknowledging that this is not the only division in the typology.

The structure of the paper is as follows: Section 2 reviews the difference between vowel intrusion and epenthesis in Articulatory Phonology. Section 3 discusses recent experimental approaches to describing inserted vowels, and the challenges these methods may pose. Section 4 lays out an expanded typology that includes subclasses of intrusive and epenthetic vowels, as well as variation between epenthesis and intrusion. Section 5 discusses work that has extended the concept of vowel intrusion to vowels inserted word-initially, word-finally, or in VC transitions.

2 Intrusion vs epenthesis

The basic difference between vowel intrusion and vowel epenthesis is depicted in Figure 1. Both start with an underlying consonant cluster. In epenthesis, a new vowel gesture is added to the representation. The new vowel acts as a syllable nucleus, changing the syllabification of the consonants, with potential knock-on effects for other aspects of phonology that refer to syllables, such as stress assignment or minimal word requirements. Intrusion, on the other hand, occurs when the two consonants remain phonologically a cluster, but their gestures are "underlapped", producing an open transition that sounds vowel-like. This open transition may enhance the perception of both consonants.

The cluster with intrusion is nearly identical in representation to a "plain" cluster. Both have the same number of gestures, segments, and syllables; they differ

only in the timing of the consonant constrictions. However, on an acoustic and perceptual level, the cluster with vowel intrusion is more similar to the cluster with vowel epenthesis. Both have a period of open vocal tract between the two consonant constrictions, which will sound vocalic. This mismatch between articulatory and perceptual characteristics is what can make vowel intrusion confusing: the vowels are "there and not there", as Ring & Gruber (2014) put it.
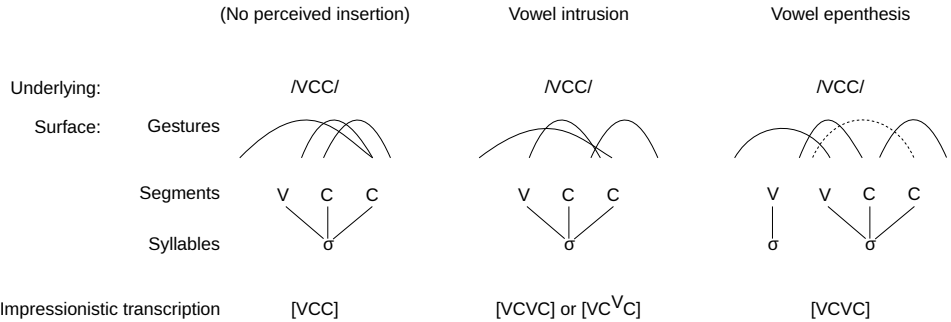

Figure 1: Gestural representations of consonant clusters without intrusion, consonant clusters with intrusive vowels, and epenthetic vowels. Dotted line = inserted gesture.

The different natures of these vocoids are reflected in different phonological patterning, different phonetic characteristics, and differences in speaker intuitions. Hall (2003, 2006) argues, based on a typological survey, that vowels which are plausibly intrusive tend to have a cluster of properties in common, as summarized in Table 1.

This clustering of properties can perhaps be best seen when both intrusive and epenthetic vowels co-occur in a single language. Two examples on which there has been recent work are Turkish and Tashlhiyt Berber.

Turkish has been described as inserting vowels into loanwords, both in CC onsets, as in [buluʒin] 'blue jeans', and CC codas, as in /sɑbr/ [sɑbɯr] 'patience'. Bellik (2019) argues that the vowels inserted in onsets are primarily intrusive, while those inserted in codas are epenthetic. The vowels in onsets are optional, and their acoustic quality is gradient and variable, with strong coarticulatory effects of nearby segments. Their durations are shorter than those of lexical vowels. They are not written in standard orthography. Bellik's ultrasound study finds that the tongue position is compatible with a "targetless" vowel. The vowels inserted in CC codas, on the other hand, seem to be true epenthetic vowels. They have a syllable repair function, occurring in marked, rising sonority coda clusters.

Table 1: Typical characteristics of intrusive and epenthetic vowels (adapted from Hall 2003, 2006)

Intrusive vowel	Epenthetic vowel
Phonetic and distributional properties	
Quality is transitional: either schwa, or influenced by surrounding consonants and vowels	Quality is grammatically determined; may be fixed or copied. A fixed-quality vowel does not have to be schwa
If the vowel copies the quality of another vowel over an intervening consonant, that consonant is a sonorant or guttural	If the vowel's quality is copied, there are not necessarily restrictions as to which consonants may be copied over
Likely to occur only in heterorganic clusters (or near taps)	Occurs in marked clusters, including homorganic ones
Likely to occur only if an adjacent consonant is voiced; voiceless clusters in the same language may have "aspiration"	Likely to have no specific association with voicing
Likely to be optional, have a highly variable duration and/or disappear at fast speech rates	Presence is not dependent on speech rate
Interaction with phonology	
May enhance perception in clusters, but does not have the function of repairing illicit clusters. May occur in relatively unmarked clusters	Repairs a marked structure, in the sense of something cross-linguistically rare, or avoided by other means in the same language
Does not form the nucleus of a syllable, cannot count for syllable-counting prosody	Forms the nucleus of a syllable
Meta-linguistic awareness	
Speakers may be unaware that they produce the vowel, and are unlikely to write it	Speakers will be aware of the vowel's presence

They are obligatory, written in the orthography, receive stress, and participate in a categorical phonological process of vowel harmony. While superficially, the insertion of vowels in onsets and codas might appear symmetrical in transcriptions, calling on the intrusion/epenthesis dichotomy helps to capture this cluster of ways in which the two phenomena differ.

A rather different example of a language that seems to show both vowel insertion types is Tashlhiyt Berber. Tashlhiyt allows long sequences of consonants, which are realized phonetically with intervening schwa-like vowels, as in /t-bdg/ [təbədəg] 'she is wet'. Since at least Dell & Elmedlaoui (1985), some linguists have argued that this schwa is a phonetic transition rather than a segment. As summarized by Ridouane & Cooper-Leavitt (2019: 434):

> Native speakers are largely unaware of the presence of this schwa in their speech, it does not affect intuitions about syllabification, it does not contribute to syllable weight in versification of traditional songs and it is ignored by phonological processes such as regressive obstruent devoicing.

It also shows distributional characteristics typical of an intrusive vowel, such as not appearing within homorganic clusters or voiceless clusters (where the transition is likely to be devoiced and hence non-vowel-like). In addition, Tashlhiyt has a longer schwa, whose insertion is prosodically triggered. Long schwa appears when segmental material is needed to allow expression of a complex final intonational melody in an emphatic statement, as in /imsʜ/ [imsəːʜ] 'he erased'. Ridouane & Cooper-Leavitt (2019) argue that the two schwas should be seen as separate elements, which they term T-vocoid (transitional) and P-vocoid (prosodic). The T-vocoid appears to be intrusive, and the P-vocoid epenthetic, albeit driven by intonation rather than syllabification.

3 Experimental approaches

Hall (2003, 2006) drew largely on impressionistic written descriptions to identify likely cases of vowel intrusion. At that time, relatively few vowel insertion phenomena had been studied in the lab, with the notable exceptions of Scots Gaelic (e.g., Ladefoged et al. 1998, Bosch & De Jong 1997) and Dutch (e.g., Van Donselaar et al. 1999, Warner et al. 2001). Since then, there has been an increase in the use of experimental methods to test the phonetic and psychological nature of inserted vowels. New studies have greatly expanded the empirical basis for analyzing vowel typology. However, interpretation of experimental results is not always straightforward, and points to a need to continue developing theories of

how abstract phonological representations relate to acoustic characteristics and speaker behaviors.

The gold standard of evidence for a gestural representation is direct imaging of articulation. A few studies have used ultrasound to test whether tongue position during an inserted vowel was consistent with a purely transitional trajectory (intrusion), or whether there was evidence of the vowel having its own target constriction (epenthesis). Davidson & Stone (2003) examined English speakers' nonce productions of pseudo-loanwords with illegal clusters, such as *zgomu* [zə-gomu], and concluded that the schwas they inserted were transitional. Bellik (2019) drew the same conclusion regarding vowels inserted in real Turkish loanwords (see also Bellik 2024 [this volume]). On the other hand, Buchwald et al. (2006) found that schwas inserted in the speech of an aphasic did show evidence of having a distinct target. Such studies unfortunately remain rare, due to the expense and difficulty of articulatory imaging.

3.1 Acoustic measures

Acoustic studies of vowel insertion are far more common, and have taken a variety of approaches to testing whether a vowel is likely to be associated with a gesture. A common finding is that putative intrusive vowels have shorter durations than lexical schwa or other unstressed vowels (see Ring & Gruber 2014 for Pnar, Garmann et al. 2021 for Norwegian, Bellik 2019 for Turkish). It is not clear that this is always true, however. For example, Tabain & Hellwig (2022: 14) find that excrescent schwa in Qaqet has a duration comparable to that of phonemic schwa, 57 vs 56 ms, as well as similar formant values. (The evidence for its excrescent status comes primarily from speaker intuitions and phonological patterning.)

As noted in Section 2, intrusive vowels in voiced clusters are predicted to correspond to voiceless transitions (often called "aspiration") in voiceless clusters. Both result from the same gestural phasing. Accordingly, some studies examine voiced and voiceless transitions together. In Khmer, for example, Butler (2015) compares words containing vowels or aspiration that she suspects to be intrusive (C_CVC, where _ is the location of a voiced or voiceless transitional element), against CʌˈCVC disyllables and CʌC monosyllables. She finds that the voicing of the transition is predictable from the voicing of the preceding consonant, and that voiced and voiceless transitions have the same average duration, which is shorter than unstressed lexical [ʌ]. The formant values of the transitional vowels are also significantly different from those of [ʌ], and more influenced by the place of the preceding consonant. She concludes that the schwa-like sounds heard in /CCVC/ words are indeed transitions within monosyllables, and that

consequently "the distinction between monosyllables and disyllables in Khmer is more clear-cut than previously thought." She also suggests that the typological concept of "sesquisyllables", as applied across languages to words with initial Cə- elements, has conflated structurally distinct word types.

In cases where it is difficult to directly compare inserted and lexical vowels in equivalent contexts, sometimes each can be compared to some point of reference in the same word. Karlin (2021) uses this technique in a small-scale corpus study of vowel insertion in Finnish archival recordings. She compares the vowels inserted in certain medial CC clusters, as in *silmä* [silimä] 'eye', against words with lexical vowels in the same position, such as *niminen* 'so-named'. To control for the variety of speakers and prosodic contexts in the corpus, the measures used were the duration ratio of V_2 (inserted or lexical) to the lexical V_1 of the same word, and the Euclidean distance between the qualities of V_2 and V_1. By both measures, vowels inserted in most /CC/ clusters patterned with lexical vowels in the same position, suggesting they are not synchronically intrusive (contra impressionistic descriptions of speaker intuitions from Harms 1976: 74 and Wiik 1965: 28). However, in one dialect the vowels in /hC/ clusters did show the short duration and variable presence that are characteristic of intrusion.

While the studies above compared inserted to lexical vowels, acoustic measures can also reveal diversity among inserted vowels. For Tripolitanian Libyan Arabic, Heselwood et al. (2015) examined inter-consonantal intervals (ICIs) in sequences of 2–4 stops, in phrases like /fak#tkasir/ 'jaw broke'. Previous descriptions had suggested that vowel epenthesis is possible in any CC sequence. The researchers find that on a range of measures, ICIs fall into two groups. Their durations have a bimodal distribution, with peaks around 20 ms and 50 ms. The shorter ICIs have the characteristics of intrusion: they tend to disappear at fast speech rates, and are often voiceless when between voiceless consonants. Short ICIs do not block voicing assimilation between flanking consonants. The longer ICIs have characteristics of epenthesis: they are more consistently present, more consistently voiced, and block consonant voicing assimilation. Short and long ICIs are generally associated with different insertion sites, although one context has both (as discussed further in section 4.3).

3.2 Challenges in interpreting acoustic results

The interpretation of acoustic results can be ambiguous, particularly when studies look beyond the inserted vowel itself to effects elsewhere in the word. Vowel insertion can be associated with durational adjustment throughout the word, not always in ways that are expected. For example, when Van Donselaar et al. (1999:

65) recorded the same Dutch words with and without schwa insertion to use as perception stimuli (e.g., *tulp* ~ *tul*[ə]*p* 'tulip'), they were surprised to find that the tokens with schwa insertion were shorter than those without. Hickey (1985) similarly claims that vowel insertion in Irish English codas cooccurs with short-ening of the preceding vowel and sonorant, so that dialectal variants like [fɪ:lm ~ fɪlɪm] *film* have the same overall duration. Gick & Wilson (2006) present similar findings for optional schwa insertion between high tense vowels and liquids in English: the rimes of words like *hee*(ə)*l* have the same duration with or without the schwa. They interpret this as evidence that the schwa is transitional.

However, it is not always clear which global durational adjustments are consis-tent with gestural retiming (intrusion) versus gesture insertion (epenthesis), nor exactly how each phenomenon would interact with other duration-affecting fac-tors such as pre-boundary lengthening, polysyllabic shortening, or speech rate variation. Interpretations typically rely on a researcher's intuitive understanding of how different gestural structures would be realized, not backed by computa-tional simulations within a spelled-out model of speech timing (like those of Gafos 2002 or Browman & Goldstein 1992a).

An example of case where interpretations might differ is Smith (2019)'s acous-tic study of schwa insertion in Scottish English codas. Smith establishes that the inserted schwas in words like *form* [forəm] are shorter than the underlying vowels in words like *forum*, averaging 54 versus 95 ms respectively, and that the inserted schwas' duration positively correlates with that of other sounds in the word. Smith interprets both facts as evidence that the vowels are epenthetic rather than intrusive. She reasons that purely transitional vowels would be less than half the duration of underlying vowels, and that

> if the inserted vowel is an articulatory byproduct of low degree of overlap between the gestural phases of the surrounding consonants, its duration should not vary with duration of the surrounding segments, but remain uniformly short (Smith 2019: 133).

To me, neither assumption is obvious. The articulatory simulations in Brow-man & Goldstein (1992a: 55) show an example of a targetless (i.e., intrusive) schwa with duration well over 50 ms, and it seems conceivable that slower speech could be associated with lower gestural overlap, producing longer transitions. I also find it interesting that Smith reports greater rates of schwa insertion in CC co-das (*farm*) than in CCC codas (*farms, farmed*). This seems opposite what one would expect for structure-repairing epenthesis, which ought to preferentially target the most marked syllables. On the intrusion analysis, on the other hand, it

seems conceivable that CCC codas are produced with greater gestural crowding and overlap than CC codas, which would decrease the chance of open transition.

However, these differences in interpretation must remain a matter of opinion until a specific hypothesis about the gestural structure of Scottish complex codas is simulated and tested against natural data. There is a great need for more articulatory simulation work to clarify what phonetic patterns can and cannot be produced through retiming alone.

3.3 Speaker intuitions and metalinguistic tasks

Since intrusive vowels are not phonological units, they are expected to be less salient to speakers' conscious awareness than regular consonants and vowels. A number of studies have probed speaker intuitions concerning inserted vowels, and frequently find them to be treated differently than lexical vowels. Yet interpretation of these differences can be difficult, due to unclarity about what level of representation various tasks draw on.

The most basic question is whether speakers are even aware the vowel is there. This is often apparently assessed through discussion or questioning, although methods for eliciting these judgments are rarely described in detail. For example, Garmann et al. (2021) report of the very short vowels found in words like [bəǀɔː] 'blue' that "anecdotally, native speakers of Norwegian tend not to be aware of the intrusive vowel, and report difficulty in perceiving it even when the articulation is pointed out to them." Similarly, Karlin (2021) directly asked Finnish speakers whether vowel insertion is possible in various /rC/ clusters. She cites their lack of awareness as evidence for the intrusive status of these vowels (in /rC/ clusters specifically, unlike other CC clusters).

Speakers can also be directly asked for syllable count judgments, but responses may be contaminated by knowledge of standard orthography as well as prescriptive notions of the syllable. Bellik (2019: 164) recounts one conversation among Turkish speakers:

> Speaker M initially said that *prens* [pirens] and *kral* [kɯral] were disyllabic. Speakers S and E then asserted that M was wrong, and *prens* and *kral* were monosyllabic. S and E argued that the definition of a syllable is a vowel, and since *prens* and *kral* are spelled with only one vowel, they were by definition monosyllabic. Faced with this argument, M lost confidence in his judgement of bisyllabicity. The fourth speaker, N, was reluctant to take a side.

Another speaker claimed that *prens* has one and a half syllables. This phenomenon of speakers counting non-whole numbers of syllables (including in

tasks where that wasn't intended to be an option) has also been reported for Scots Gaelic (Hammond et al. 2014), and for English words like *fire* (Cohn & Tilsen 2015). These stories suggest that speakers' definition of a "syllable" does not always align with the concept of a syllable in phonological theory. It is important to design instruments, like that of Cohn & Tilsen, that allow speakers to give intermediate responses, as that detail will be lost if they are forced into whole-syllable judgments. The question of what exactly speakers are counting when they talk about half syllables remains unclear. It could be a way of describing a non-syllabic phonetic sonority peak, such as an intrusive vowel.

Some studies, rather than asking directly, use tasks designed to indirectly access syllabicity judgments. Ring & Gruber (2014) used a language game (based on a method developed by Van Donselaar et al. 1999) to test the extra-short vowels that Pnar speakers optionally insert in onset clusters, such as [kba ~ kəba] 'rice'. Speakers heard audio recordings of real and pseudo words. They were told to reverse the phonemes in any word that was a monosyllable (tam → mat), but reverse the syllables of disyllables (batam -> tamba). Both real and pseudo words with inserted schwa were predominantly treated as monosyllables. Real words with lexical schwa in the first syllable were treated as disyllables, even when the schwa was clipped to the typical duration of an inserted vowel. This suggests that speakers have different representations of lexical and non-lexical schwa, and do not consider the inserted schwa syllabic, consistent with the predictions of the intrusion analysis.

In Turkish, Bellik (2019) examines whether intrusive and lexical vowels are treated alike in text-to-tune setting. In a corpus of recorded songs, she finds that vowels inserted in complex onsets can be set to a beat (48/82 tokens), but often are not (34/82), while lexical vowels in the same position are almost always set to at least one beat. There is variation across lexical items, where the words with the most consistent insertion are most likely to have a beat for the inserted vowel. However, an elicitation experiment had different results. When asked to insert words into a two-note slot, speakers either suppressed vowel insertion, or produced it and gave the intrusive vowel its own note. This is problematic for the intrusion analysis, given that the presence/absence of an inserted vowel changed the metrical setting (but see discussion in section 4.3). See also Baronian & Royer-Artuso (2024 [this volume]) for discussion of issues in interpreting text-setting as evidence in phonology.

Although standardized orthography can contaminate judgments, spontaneous or unstandardized orthography can be a useful indicator of speaker intuitions. For example, Qaqet has optional schwas, mostly adjacent to liquids as in [mraɽik ~ məraɽik] 'cross', as well as a phonemic schwa as in [aləm] 'feather' (Tabain &

Hellwig 2022). The vowels differ in phonological behavior, with only the phonemic schwa triggering lenition of voiceless plosives. Tabain & Hellwig (2022) claim that the optional schwas are excrescent rather than epenthetic, since they do not break illegal consonant clusters. Interestingly, they report that the two schwas are acoustically comparable in duration and formant values. However, speakers treat the two schwas differently:

> Speakers agree on the presence of the phonemic vowel when asked for an orthographic representation, but do not write the excrescent vowel, whose presence is indeed highly variable across speakers and across lexical items. (Tabain & Hellwig 2022: 14)

Other examples where speakers do not write plausibly transitional vowels include Lamkang (Burke et al. 2019) and Kalam (Blevins & Pawley 2010, discussed further in section 4.4).

A recurrent challenge in interpreting metalinguistic tasks is how to understand gradient results. For example, Hammond et al. (2014) used a battery of tasks to probe the intuitions of elderly Scots Gaelic speakers. Words with inserted vowels, such as *ainm* [ɛnɛm] 'name', were compared to words with only lexical vowels, such as *anam* [anam] 'soul'. Tests included identifying which of two words an audio (V)CVC sequence was clipped from, writing nonce words in Gaelic orthography, counting syllables, knocking syllables, and identifying the first and last syllable of the word. The syllable-identification task was intended to probe syllabification of the medial consonant, which has been reported to act as a coda before lexical vowels but an onset before inserted vowels. Speakers treated inserted vowels differently than lexical vowels on nearly every measure, but in every case showed intermediate and inconsistent results. For example, speakers counted the inserted vowel as a syllable 56% and 59% of the time in the counting and knocking tasks, as opposed to 82% and 87% for lexical vowels, and a number of participants tried to offer answers such as "one and a half syllables". Syllabification results were also puzzling: when asked to identify the first and second syllables of CVCVC words with inserted vowels, speakers tended to not to produce the medial consonant in either syllable, leaving its structural affiliation unclear. The authors conclude that the inserted vowels are phonological objects and not phonetic transitions, given that they affect syllabification and count as syllables at least some of the time for speaker intuitions.

In some cases, it can be debatable what level of linguistic structure a given speaker behavior taps into. For example, Burke et al. (2019) describe an ambiguous case in the Trans-Himalayan language Lamkang. Lamkang optionally

inserts extra-short vowels into strings of consonants. Up to 4 consonants can occur word-initially, through prefixing, as in *mtknoolam* 'they are shaking me off.' Phonetically, such a word can be realized as [mᵊtᵊkᵊnoolɐm], with extra-short vowels between the concatenated consonants. Extra-short vowels average 33 ms in duration, compared to 85 ms for lexical short vowels. Although missionary translators wrote the vowels in the orthography, native speakers do not. These acoustic and orthographic characteristics seem consistent with the vowels being intrusive.

Yet when Burke et al. (2019) asked speakers to clap or tap to the syllables in such words, they clapped to each extra-short vowel. Variation in vowel insertion corresponded with variation in clapping. Among 6 speakers there were 4 pronunciations of the word /m-t-pmen/ 'she is trapping me', with 2–4 claps: [mᵊtpmɛn], [mᵊttᵊpmɛn], [mᵊtpᵊmɛn], and [mᵊttᵊppᵊmɛn]. Burke et al conclude from the clapping task that the extra-short vowel does act as a syllable nucleus, which would mean it is not intrusive.

However, the high variability in the responses seems atypical for syllable count judgments. This raises the question of whether the clapping task is actually tapping into intuitions about syllables in the phonologist's sense, or something else. Clapping is known to reflect different structures in different languages; while English speakers clap to syllables, Japanese speakers clap to moras. It could be that clapping in Lamkang reflects surface intensity peaks, or that it taps into some type of timing structure that is used in coordinating long series of consonants.

In short, metalinguistic tasks are turning up fascinating data on speakers' knowledge of lexical and inserted vowels, which require explanation. Yet interpretation of the facts can be difficult. There isn't an explicit theory of how each metalinguistic task relates to phonological representations in each language. We don't always know to what extent different tasks tap into knowledge of orthographic forms, underlying lexical forms, surface phonological representations, and/or purely acoustic forms. Methodological refinements may clarify some cases where a metalinguistic behavior seems to not match other indicators of transitional vs segmental status, such as Lamkang.

3.4 Orthography and historical development

Phonological representations can influence spelling, but Bellik (2019) has suggested that influence may also flow the other way: an orthographic convention of not writing a vowel could potentially reinforce speakers' internal representation of the vowel as transitional, delaying the common diachronic path whereby intrusive vowels are eventually phonologized as segments.

Bellik (2019) points out that vowels in older Turkish loans seem to have been reanalyzed as segments and lexicalized fairly rapidly. Walter (2018) shows that in Ottoman-era texts, many loans from European languages were written down with an inserted vowel, and those vowels are lexicalized today. Yet Bellik (2019) argues that newer loans seem to be retaining vowel intrusion as a somewhat stable configuration. She suggests that the difference relates to speakers' metalinguistic knowledge. A younger generation of the Turkish speakers are aware of the prescriptive spelling of newer loans, with no orthographic vowel corresponding to the intrusive vowel. They are also likely to have studied English or other source languages for the loanwords, and hence be aware that the source words have a consonant cluster. Both factors increase the likelihood that they will continue to interpret the vocalic period in CC onsets as an open transition rather than a vowel segment. Bellik cites anecdotal evidence that children often include the vowel when learning to write, and are taught not to. This suggests that intrusive vowels in these loanwords would naturally tend to be on their way to reanalysis as segments, were it not for external factors.

4 Expanding the typology

As discussed in Section 3, in-depth examination of any specific vowel insertion process by multiple measures tends to find less-than-perfect adherence to the list of prototypical qualities of intrusive and epenthetic vowels given in Table 1. Some linguists have concluded from this that, in the words of Hammond et al. (2014), "any phonology-phonetics distinction for inserted vowels is probably more of a continuum, rather than a categorical split, or at least that there are more than two types of inserted vowels." In this section I lay out one version of what a more articulated typology could look like. This proposal retains the basic dichotomy of presence/absence of a vowel gesture, but acknowledges three types of additional phonetic variation: 1) the existence of more than one gestural timing configuration that can produce vowel intrusion, 2) the existence of gesturally stronger and weaker types of epenthetic vowel, and 3) the possibility that what linguists describe as one vowel insertion process sometimes involves a mixture of epenthesis and intrusion. This typology offers more ways to categorize inserted vowels, while keeping the insight that a definite set of gestural facts need to underlie each type. The predictions of this theory remain more restrictive than appeal to a loosely-defined continuum.

Note that this typology does not attempt to address purely phonological diversity among true epenthetic vowels, such as variation in their quality, contexts,

interaction with stress, etc. I assume that these characteristics are controlled by phonological constraints of some type, and are not limited by the vowels' phonetic substance, which is the focus here.

4.1 Subtypes of vowel intrusion

Intrusive vowels are not all the same, because they can result from different levels of consonant overlap. This was established by Gafos (2002), whose simulation work showed that some characteristics of the intrusive vowel are expected to change as the overlap decreases, as depicted in Figure 2 and (1). At a moderate degree of overlap (when the articulation of C_2 starts around the articulatory midpoint of C_1), transitions within consonant clusters will occur only at slower speech rates and in heterorganic clusters. Consonant voicing also has a strong influence at this level of overlap; the transition will be voiceless (often described as "aspiration") if the flanking consonants are voiceless. But all of these characteristics change if the consonants have little or no overlap, as shown on the right. In clusters of non-overlapping consonants, release will also occur even if they are homorganic, even at faster speech rates, and even in voiceless environments.
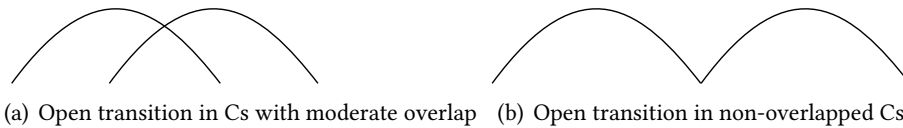

(a) Open transition in Cs with moderate overlap   (b) Open transition in non-overlapped Cs

Figure 2: Subtypes of vowel intrusion

(1)   Subtypes of vowel intrusion

   a. Open transition in Cs with moderate overlap
   - Disappears in fast speech
   - Occurs in heterorganic clusters only
   - Transition is voiced only near voiced C; realized as "aspiration" in voiceless environment
   - Example: Sierra Popoluca (Gafos 2002): [miɲᵊpaʔ]

   b. Open transition in non-overlapped Cs
   - Present in fast speech
   - May occur in homorganic clusters
   - May be voiced between voiceless Cs
   - Example: Moroccan Colloquial Arabic (Gafos 2002): [zˤnatˤᵊtˤ]

It is predicted that we should find some cases of vowel intrusion that occur in a wider variety of cluster types, and are less speech-rate dependent than others. This is one reason that cases of genuine intrusion could depart from the prototypical list of characteristics given in Table 1, and why certain characteristics in that table are hedged as "likely", rather than required. If a vowel has several characteristics of intrusion but not exactly the typical distribution pattern, it is worth considering that it could be the lower-overlap type.

4.2 Phonetic subtypes of epenthetic vowels

A second source of variation among inserted vowels is that when a new gesture is inserted, it is not necessarily always the same kind of gesture, or phased the same way with respect to other gestures. There is evidence for phonetic diversity among true epenthetic vowels. Some appear to be phonetically identical to lexical vowels in the same language, while others are not.

For example, Hall (2013) finds that some (but not all) Lebanese Arabic speakers produce epenthetic vowels that are shorter in duration and more centralized in quality than lexical vowels. Words like [libis] 'wore' (from /libis/) and [libis] 'clothes' (from /libs/) are not quite homophonous. Yet the inserted vowels are not merely transitional; they have all the characteristics of true epenthesis. They have a fixed quality, which varies regionally from [i] to [e] to [ə], and they can have a long duration in certain prosodic contexts such as phrase-final lengthening. Their presence is optional, but in an all-or-nothing way rather than a phonetically gradient way. Vowel insertion occurs most frequently in rising sonority CC codas, where it has a clear syllable-repair function. It can occur in homorganic clusters like /nt/, which are less prone to open transition. Speakers are consciously aware of the vowels, will write them if asked to write in the colloquial, and can suppress them when speaking in a formal register. All of this suggests that the inserted vowels have their own gestures, which are not necessarily identical to the gestures of similar lexical vowels.

I am not aware of any explicit proposal for how to model phonetically weak true epenthetic vowels, like those in Lebanese, in gestural representations. Weak epenthetic gestures could have special settings for properties such as stiffness, which affects duration, or blending strength, which affects how much two competing gestures influence an articulator's trajectory. Another possibility is that epenthetic vowels are more overlapped by nearby consonants than lexical vowels are. Some hypothetical gestural representations of weak epenthetic vowels are shown on the left of Figure 3.

(a) Epenthetic V with short duration

(b) Epenthetic V with heavy overlap
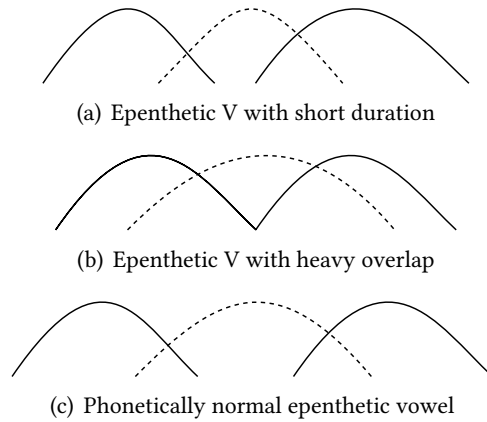

(c) Phonetically normal epenthetic vowel

Figure 3: Phonetic subtypes of epenthetic vowels. Hypothetical representations of phonetically weak and strong epenthetic vowels. (Dotted line = inserted vowel)

Epenthetic vowels are not always phonetically weak, however. Hall (2013) found that some Lebanese speakers have epenthetic vowels that are acoustically indistinguishable from lexical [i]. Other reports of phonetically normal epenthetic vowels include Mohawk (Michelson 1989: 40) and vowels in Korean loanwords (Kim & Kochetov 2011). This situation is depicted on the right of Figure 3.

Phonetic weakness in true epenthetic vowels could be a natural and fleeting stage of diachronic development, or it might persist for functional reasons. Epenthetic vowels carry no lexical information of their own. Keeping them phonetically reduced is in line with a general tendency to maximize informative material and minimize predictable material. It would be useful to have more studies of the phonetics of true epenthetic vowels, to better understand the range of variation within this category.

4.3 Variation between gesture presence and absence

A final complication for classifying vowels as intrusive or epenthetic is that there could be variation between gestural presence and absence. This possibility is depicted in Figure 4. Such variation could occur across phonological contexts, or across speakers, or across registers.

Alternation between an intrusive vowel and a segmental vowel is a logical possibility anywhere that segmental CC ~ CVC variation occurs. One piece of evidence that this actually does occur is Heselwood et al. (2015)'s acoustic study

$$[\text{CəC}] \quad \sim \quad [\text{CəC}]$$

Figure 4: Inconsistent presence of an epenthetic vowel gesture, with open transition (vowel intrusion) in cluster when vowel gesture is absent. Example: Tripolitanian Libyan Arabic (Heselwood et al. 2015): [hatˀk ~hatək].

of Tripolitanian Libyan Arabic. As noted in section 3.1, they find a bimodal distribution of durations for inter-consonantal intervals (ICIs) in lexical consonant clusters, with short ICIs apparently intrusive and longer ICIs apparently epenthetic. Both epenthesis and intrusion occur in phrase-final consonant clusters, as in [hatˀk ~ hatək] 'violation' (transcription inferred from the source). The distribution of ICIs in this specific context shows separate peaks above and below the typical threshold for epenthesis in the dataset. The epenthesis/intrusion distinction can explain why the distribution is specifically bimodal, rather than a wide normal distribution. Note that this level of detail would probably be missed, at least by non-native speakers, without such an acoustic study.

Intrusion/epenthesis variation offers a tool for analyzing certain problematic cases. Another situation where I suspect this might be useful is explaining one anomalous result from Bellik (2019)'s study of intrusion in Turkish CC onsets. As discussed in section 3.3, there is acoustic and articulatory evidence that these vowels are often transitional in speech. Yet when singing such words as part of a text-setting task, speakers who produced the inserted vowel nearly always gave it a note, suggesting it is epenthetic rather than intrusive in that circumstance. One way to reconcile these findings is to consider that speakers might be producing different vowel types in different contexts. Bellik (2019: 179) notes of the singing task that:

> Due to the nature of the task, the non-lexical vowels were produced in a way that sounded impressionistically like a full vowel, in contrast to the brief and schwa-like vowels heard in the acoustic and ultrasound production experiment.

This suggests that some speakers produce intrusive vowels in casual speech, but epenthesize a full vowel gesture as a stylistic flourish when singing. This is somewhat analogous to the P-vocoids that Ridouane & Cooper-Leavitt (2019) posit for Tashlhiyt Berber: epenthetic vowels which occur specifically when needed for realization of a prosodic tune.

Variation between gestural presence and absence is likely to occur at some stage during a diachronic process of reanalysis. Hall (2006) notes a number of cases where vowels that likely began as intrusive are now segmental. But reanalysis does not happen all at once. We would expect a transitional period in which the gestural structures vary, certainly across speakers and possibly within speakers as well. E.g., a single speaker might produce a consonant cluster sometimes with a phonetic transition and other times with an inserted vowel gesture, possibly depending on speech register or prosodic factors. Some innovative speakers might reanalyze all intrusive vowels as segments, at a stage where other speakers still produce them as intrusive.

This possibility of variation between intrusive and segmental vowels obviously complicates the task of describing and classifying vowels. However, recognizing this possibility can help disentangle cases where a vowel insertion pattern seems to be inconsistent.

4.4 Mixed behavior of remnant vowels

Another case where a mix of segmental and nonsegmental vowels might plausibly occur is in what Blevins & Pawley (2010) call the "remnant vowels" of Kalam. Remnant vowels stem from a diachronic reduction process, in which former full vowels reduce to schwa, and then possibly to a purely transitional schwa. Blevins & Pawley (2010) show that that in Kalam, remnant vowels display some properties associated with intrusive vowels and other properties associated with epenthetic vowels[2]. They consider this mixture of characteristics problematic for classifying the vowels as epenthetic or intrusive. However, it is possible that the class of remnant vowels actually encompasses both types.

The vowels in question are also known as "predictable vowels", and occur between all adjacent consonant phonemes, as in /pk-p-n-p/ [ɸɨɣɨβɨnɨp] 'I could have hit' (Blevins & Pawley 2010: 20). Their quality is either central or influenced by nearby consonants or vowels. Speakers do not write them. The vowels appear in contexts where they are not required for repair; they superficially appear to create CV syllables although CVC is licit in the language. All of these characteristics are very typical of transitional, non-segmental sounds.

A couple of other characteristics are less typical of intrusion, but not impossible. The predictable vowels are not speech-rate dependent, and can appear in homorganic clusters. As discussed in section 4.1, these characteristics are predicted to occur when there is a little or no CC overlap (see Figure 2).

---

[2]Huang (2018) analyzes a similar case in Squliq Atayal (Formosan), which I will not describe here for space, but it poses similar issues.

However, certain predictable vowels show two phonological behaviors that are inconsistent with the phonological invisibility expected of intrusive vowels. First, predictable vowels can appear word-finally to 'bulk up' a subminimal word that would otherwise be just a consonant, as in /m/ [mə] 'taro'. Enforcing word minimality is a phonological repair function typical of epenthetic vowels. Also, predictable vowels bear stress if they are the final vowel in the word, as in /mlp/ [mɨˈlɨp] 'dry'. Both facts are strong evidence for presence of a phonological unit, not merely a transition.

It might be possible to capture both aspects of the vowels' patterning by analyzing predictable vowels as consisting of two distinct groups, similar to Ridouane & Cooper-Leavitt (2019)'s claim that there are both transitional and non-transitional schwas in Tashlhiyt Berber. Although Kalam predictable vowels all result from a diachronic reduction, it could be that reduction was complete in some contexts and incomplete in others. Perhaps in unstressed position, the language experienced full loss of etymological vowels, including loss of their gestures, leaving behind consonant clusters that were phased with very low overlap so as to continue to produce open transitions. In stressed position and in minimal words, however, full deletion was blocked for phonological reasons. Deletion could not occur if the vowel segment was needed to realize stress, or to keep a minimal amount of phonological material in a word. In these contexts, reduction resulted in a vowel that lost its original quality but still had a gesture and formed a syllable. On this analysis, some Kalam predictable vowels are intrusive while others are segmental.

A challenge for this approach is to explain why speakers *don't* write the predictable vowels in cases where they are needed for word minimality, such as /m/ [mə] 'taro'. This does not fit with the assumption that speakers typically write segments, including epenthetic segments. It is possible that such spellings reflect underlying representations, which surface without the predictable vowel in related words like /m-adeŋ/ [maˑndeˑŋ] 'taro plant sp.'

4.5 Symmetrical vowel intrusion? A reconsideration of Hoocąk

Hall (2003) argues that Scots Gaelic and Hoocąk (Winnebago) have a special type of "symmetrical" vowel intrusion. Both languages have inserted vowels that are relatively long and distinct in quality, yet have a number of characteristics usually associated with intrusion. Briefly, the theoretical proposal is that a sonorant and vowel are adjoined to create a bimoraic nucleus, similar to a diphthong. This is different from most vowel intrusion processes, which typically involve a transition within a complex onset or coda. In this diphthong-like structure, the vowel

and sonorant are unordered with respect to one another, and articulatorily realized with their centers aligned. The vowel gesture fully envelops the sonorant, so that periods of the vowel are heard both before and after the sonorant. The overall duration of this bimoraic nucleus is like that of a long vowel.
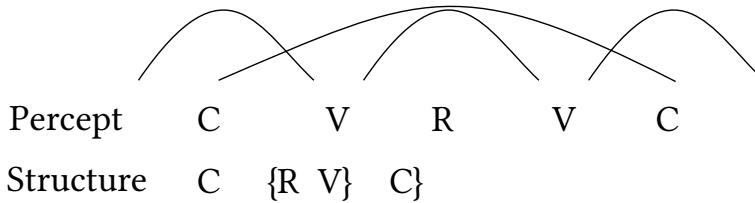

Figure 5: Symmetrical intrusive vowels. C = consonant, R = sonorant, V = vowel. R and V are phonologically unordered with respect to one another.

This proposal attempts to explain why Scots Gaelic and Hoocąk have CVRVC sequences with unusual phonetic properties that pattern with monosyllables in a variety of ways (see Hall 2003, chapters 4 and 5 for details). However, I have come to doubt that this is the correct synchronic structure of inserted vowels in Hoocąk.

For background, Hoocąk inserts copy vowels in underlying /ÇRV/ through "Dorsey's Law", as in /sni/ [sĩnĩ] 'cold'. The resulting [ÇV$_i$RV$_i$] structures, known as Dorsey's Law sequences, differ from typical disyllables phonetically and phonologically. Steriade et al. (1990) proposed that the two vowels could be a single gesture, while Alderete (1995) proposed that the sequences are parsed as monosyllables. The symmetrical intrusion analysis ties these ideas.

One problem for this analysis comes from a phonological alternation I had previously overlooked. Susman (1943) reports that the lexical (second) vowel of a Dorsey's Law sequence can undergo raising independently of the inserted vowel. This occurs before suffixes beginning with /-(h)a/, as in /kɾe + hak/ → [keɾiak] 'he barked moving' (Susman 1943: 75). The second vowel can also undergo independent lengthening before the declarative suffix /-ną/, as in [kiɾi:ną] 'he returned' (Susman 1943: 14). In both cases, it looks like the second vowel must be a separate gesture from the first.

Recently, a group in my lab has annotated and begun to analyze archival recordings of a speaker who worked with Kenneth Miner in the 1970's (Hall et al. 2019). We find that Dorsey's Law sequences do have some characteristics that could be consistent with the symmetrical intrusion analysis: they are audibly unlike regular disyllables, with shorter overall duration, and the speaker seems

to always produce them as a rapid unit; we have not found cases of pauses or disfluencies within the sequence.

On the other hand, we do not find support for a much-cited claim that the inserted vowels have special accentual properties, which had been important evidence for the intrusion account. Miner (1979) reported that while normal disyllables had one accent, as in [xatáp] 'brush', Dorsey's Law sequences were accented on both vowels, as in [kèɾé] 'depart returning'. This apparently shared accent seemed like evidence that the inserted vowel formed a unique prosodic unit with the following vowel, compatible with the idea that the two vowels were a single syllable. However, recent work in our lab by Cameron Duval suggests that there is actually no special accent on the inserted vowel. The speaker does produce inserted vowels with a slightly higher pitch than the average lexical vowel in the same position, but the effect is small, and plausibly attributable to the fact that inserted vowels always follow voiceless obstruents, which cross-linguistically raise the pitch of following vowels.

In short, it seems likely that Hoocąk inserted vowels do synchronically have their own gesture, although its short duration indicates that it is phonetically different from lexical vowels. The reanalysis of this example, as well as Hammond et al. (2014)'s challenges to the Scots Gaelic example, make it doubtful that symmetrical vowel intrusion exists.

5 Extending the theory: Gesture-less vowels in other contexts

Although Hall (2006) focused only on intrusive vowels occurring between consonants, the concept of vowel intrusion has also been extended to vowels inserted in other contexts. Various studies have examined the likelihood of non-segmental vocalic periods arising in the transition between a vowel and consonant, or following a word-final consonant, or preceding a word-initial consonant. The evidence for each is reviewed below.

5.1 V-to-C transitions

Several studies have argued that certain gestural timing in VC sequences can produce acoustic intervals that sound like a distinct vowel, without insertion of a gesture. An example is the [a] in Tiberian Hebrew /ruħ/ > [ruaħ] 'spirit'. Operstein (2010) terms this phenomenon "consonant prevocalization", and notes that it is often below the level of speakers' linguistic consciousness, similar to the

way that speakers are often unaware of vowel-like releases between consonants. She argues, in an Articulatory Phonology framework, that these vowel-like percepts result from staggered timing of multiple gestures involved in consonants. Each consonant segment, in Operstein's approach, has both a C-gesture and a V-gesture, and sliding the V-gesture earlier produces prevocalization. The effect is the percept of a vowel, which is initially subphonemic. At this stage, the acoustic VVC sequence is in some ways analogous to the acoustic CVC sequences arising from low gestural overlap: there is a percept of an additional vowel, which a linguist can hear but a speaker may not notice, and which does not initially count as an additional syllable in the phonology. It is not truly gestureless, though, since it is produced by a distinct V-gesture (albeit one not associated with a vowel segment).

Others have argued that an intrusive vowel can occur in a V_C context without Operstein's sliding V-gesture, but rather simply due to the tongue's transition between conflicting vowel and consonant articulations. For example, in American English, a schwa-sound often occurs between high tense vowels and liquids, in words like *heel* [hiəl] or *fire* [faɪəɹ]. Gick & Wilson (2006) present evidence that this is an incidental acoustic result of the tongue moving through schwa-like configuration during its transition between conflicting targets. They find that the presence of this schwa does not lengthen the syllable it occurs in. Cohn & Tilsen (2015) find that speakers vary in their syllable count judgments for such words, particularly in the case of /aɪɹ/ and /aɪl/ rimes, as in *fire* and *file*. Words that were judged to have more than one syllable (participants were allowed to give intermediate judgments such as 1.5 syllables) were produced with longer rimes. This might suggest that such words are undergoing reanalysis, with some speakers reinterpreting the transitional schwa as a distinct gesture that forms the nucleus of a new syllable.

Similarly, Garellek (2020) argues that a process in Central Yiddish that inserts schwas between long vowels or diphthongs and certain codas, as in /biːχ/ → [biːəχ] 'book', originated as phonetic transitions. He shows that these schwas occur in contexts where the tongue would be most likely to pass through a schwa-like space when transitioning between the nucleus and coda. Garellek suggests that for some speakers, the schwas are still non-segmental, as shown by their variability. However, some 19th century poetry shows that they counted as metrically present. Like the English schwas discussed above, the Yiddish schwas seem to be prone to reinterpretation as segments.

In short, some vowel-like transitions within V_C do seem to have intrusion-like properties, and also pose similar challenges as to how to interpret gradient or ambiguous speaker intuitions, and mixed or variable patterning.

5.2 Word-final vowel insertion (paragoge)

Could word-final inserted vowels be non-segmental? Among linguists who believe that phonological processes have phonetic precursors, there is widespread agreement that final vowels arise historically as reinterpretation of audible consonant releases (Kang 2003, Blevins 2004:146, 156, Ng 2013). This may explain why, despite being rare as a regular diachronic sound change, final vowel insertion is highly common in second language speech (interlanguage), in loanword adaptation, and in languages that arise from language contact (creoles). Ng (2013) argues that these contexts have in common the slow and effortful speech of adult learners, which is more likely to involve hyper-articulated final consonants with audible releases.

But an origin in audible release does not necessarily mean that there is an intermediate stage in which gestural timing alone produces a stable vowel-like percept, analogous to intrusion. The appearance of the vowel could instead take place in a single leap (released C > CV), with a vowel gesture and segment inserted to satisfy some phonological requirement.

For example, Korean often inserts final vowels into English loans, as in *pad* [pʰæ.tɨ]. Kang (2003) points out that this process is not motivated by phonological repair, as codas like [t] are licit in Korean. Her corpus study shows that final vowel insertion is most common in exactly the contexts where English tends to have the most frequent phonetic release, such as final voiced stops. Yet, accepting that the English release is in some sense the source of the inserted vowel, there is no reason to think the inserted vowels are non-syllabic. Kim & Kochetov (2011) show that the insertion is categorical, resulting in a vowel segment whose duration and quality are comparable to those of native vowels. In this case, then, there seems to be a direct leap from the English word with audible (but non-vowel-like) release, to the Korean adaptation which inserts a full vowel in order to maintain a kind of faithfulness to the original release. There is no evidence for a progression of C > Cə > CV, with an intrusion-like stage.

One case where some linguists have argued for non-segmental final vowels is Italian. Consonant-final loanwords are optionally adapted with final schwas, along with lengthening of the final consonant, as in *jet* [dʒɛt ~ dʒɛt:ə]. Miatto (2020) finds that when asked to count syllables in recorded nonce words, Italian speakers judge tokens like [vik:ə] as monosyllabic 93% of the time. When asked to repeat the words, they sometimes insert final [ə] when it was absent from the stimulus, or leave it out when it was present. Repetti (2012) similarly argues that the final vowel "does not play a phonological role". She points to its variable phonetic quality, which she transcribed as [ə], [ᵊ], or [e] (but not identical to lexical [e]), and reports that

anecdotally, native speakers of Italian do not perceive a final vowel in consonant-final words when they are uttered by themselves or by other native speakers of Italian; however, native speakers of English often do hear a vowel in consonant-final words uttered by Italian native speakers (Repetti 2012: 175).

Grice et al. (2018) mention that there is debate over whether these vowels should appear in the orthography. Grice et al. (2018) argue that Italian final schwa shows a mixture of properties of epenthesis and intrusion. Like epenthesis, it repairs a marked structure, since consonant-final words are rare in the native vocabulary. It is variable, a common characteristic of intrusive vowels, but Grice et al. show that the variability is phonologically conditioned: schwa is more likely to appear on monosyllables than disyllables, and more likely to appear when the word has a complex intonational melody in need of segmental material to realize it. This prosodic conditioning could indicate that the schwa is a phonological element inserted for metrical reasons, similar to the P-vocoid that Ridouane & Cooper-Leavitt (2019) propose for Tashlhiyt (see Section 2). Overall, the question of whether Italian final schwa could be intrusive seems unsettled.

As far as I am aware, there are no proposed articulatory models of final intrusive vowels, where gestural alignment alone can produce a vowel-like percept in final position. One question is how the final period would become voiced, especially if the final C is voiceless as in *jet* [dʒɛt:ə]. This is not necessarily problematic, given that Articulatory Phonology treats voicing as a default state. In Browman & Goldstein (1992a), for example, there are no voicing gestures, only wide-glottis gestures to produce periods of voicelessness. In simulations, voicing can appear even in the transition between two voiceless consonants if they are long enough. In the case of final vowel insertion, if the speech system stays "on" after the end of the final oral constriction, the result could conceivably be a short period of voicing that does not correspond to any oral gesture. Whether this actually occurs is an open question. See also Hamann & Miatto (2024 [this volume]) for discussion of the language-specific perception of release bursts.

5.3 Initial vowel insertion (prothesis)

Another context in which vowels are often inserted is word-initially, typically before a CC cluster. Cross-linguistically, prothesis tends to be most common before obstruents and especially sibilants (Fleischhacker 2005). It is unclear whether some initial inserted vowels could lack a gesture. Some impressionistic descriptions seem consistent with this idea; for example, Hewson (1986: 138) notes of Mi'kmaw:

> there is normally a slight prothetic vowel before the initial /k/ in /kti/ 'thy dog' that has traditionally never been written by native speakers, who intuitively recognize it as a purely phonetic element that helps to make the cluster pronounceable.

Operstein (2010) references similar examples from grammars of Welsh and Western Armenian.

Linguists who have explored the possibility of initial intrusive vowels differ in the proposed gestural structure that would underlie such a phenomenon. Bellik (2019: 260) suggests that initial vowel insertion could occur if the nucleus vowel fully overlaps the onset consonant. Operstein (2010: 154–155) views initial vowel insertion as part of the same "prevocalization" phenomenon that also occurs in rimes, where she posits that the "V-gesture" and "C-gesture" of a single coda consonant become temporally disassociated.

Both of these proposals assume that some kind of vocalic constriction gesture must extend before the initial consonant constriction to produce the percept of an initial vowel. Another possible approach, similar to that suggested for final vowel insertion above, would be to explore ways that the default voicing associated with speech in Articulatory Phonology might turn on before the first consonant constriction, producing an initial voiced period unassociated with any oral constriction.

6 Conclusion

The description of vowel insertion processes has benefited enormously from the many careful phonetic and phonological studies reviewed here (as well as others omitted for space). There are now far more acoustic phonetic descriptions and even imaging work on inserted vowels, and more descriptions of how speakers treat the inserted vowels in spontaneous spelling, language games, text-to-tune setting, clapping tasks, etc. Some of this work has supported the intrusion/ epenthesis dichotomy, some has challenged it, and all of it lays the groundwork for more detailed and accurate typologies.

This chapter acknowledges that the distinction between vowels with and without gestures is not always as clear-cut as it seemed in Hall (2003, 2006). Characteristics that clustered in a typological survey based mostly on impressionistic descriptions (Figure 1) do not always align perfectly when a single language is probed in depth. I have argued, however, that apparent exceptions can often be explained by 1) recognizing the diversity of gestural configurations that can

produce vowel intrusion, 2) exploring the possibility of variation between transitions and gestures, and 3) acknowledging some limitations in our understanding of current methods. Methodological advances in two areas would be particularly useful: more simulations to determine what types of acoustic phenomena can be produced through gestural phrasing alone, and better understanding of what is happening in various metalinguistic judgments and tasks intended to probe speakers' phonological representations.

References

Adiego, Ignasi-Xavier. 2019. Consonant clusters, defective notation of vowels and syllable structure in Caromemphite. In Ronald I. Kim, Jana Mynářová & Peter Pavúk (eds.), *Hroznỳ and Hittite: The first hundred years* (Culture and History of the Ancient Near East 107), 95–119. Leiden: Brill. DOI: 10.1163/9789004413122_008.

Alderete, John. 1995. Winnebago accent and Dorsey's Law. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Winnebago accent and Dorsey's Law*, 21–51. Amherst: GLSA, University of Massachusetts.

Baronian, Luc & Nicolas Royer-Artuso. 2024. Textsetting the case for epenthesis in Armenian. In Ji Yea Kim, Veronica Miatto, Andrija Petrović & Lori Repetti (eds.), *Epenthesis and beyond: Recent approaches to insertion in phonology and its interfaces*, 271–289. Berlin: Language Science Press. DOI: 10.5281/zenodo.14264550.

Bellik, Jennifer. 2019. *Vowel intrusion in Turkish onset clusters*. University of California. (Doctoral dissertation).

Bellik, Jennifer. 2024. Gestural characteristics of vowel intrusion in Turkish onset clusters: An ultrasound study. In Ji Yea Kim, Veronica Miatto, Andrija Petrović & Lori Repetti (eds.), *Epenthesis and beyond: Recent approaches to insertion in phonology and its interfaces*, 143–166. Berlin: Language Science Press. DOI: 10.5281/zenodo.14264540.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette & Andrew Pawley. 2010. Typological implications of Kalam predictable vowels. *Phonology* 27(1). 1–44.

Bosch, Anna & Kenneth De Jong. 1997. The prosody of Barra Gaelic epenthetic vowels. *Studies in the Linguistic Sciences* 27(1). 1–15.

Browman, Catherine P. & Louis Goldstein. 1992a. "Targetless" schwa: An articulatory analysis. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in laboratory phonology II: Gesture, segment, prosody*, 26–67. Cambridge, UK: Cambridge University Press.

Browman, Catherine P. & Louis Goldstein. 1992b. Articulatory phonology: An overview. *Phonetica* 49(3-4). 155–180.

Buchwald, Adam, Brenda Rapp & Maureen Stone. 2006. Evidence for discrete phonological representations in production: Ultrasound imaging of aphasic speech. *Brain and Language* 99(1-2). 140–141.

Buckley, Eugene. 2007. Vowel–sonorant metathesis in Alsea. *International Journal of American Linguistics* 73(1). 1–39.

Burke, Mary, Shobhana Lakshmi Chelliah & Melissa Robinson. 2019. Excrescent vowels in Lamkang prefix sequences. *Journal of South Asian Languages and Linguistics* 6(2). 185–213.

Butler, Becky. 2015. Approaching a phonological understanding of the sesquisyllable with phonetic evidence from Khmer and Bunong. In N.J. Enfield und Bernard Comrie (ed.), *Languages of mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 443–499. Berlin: De Gruyter Mouton.

Canalis, Stefano. 2007. To what extent is Albanian word stress predictable? In Antonietta Bisetto & Francesco Barbieri (eds.), *Proceedings of the XXXIII incontro di grammatica generativa*, 1–14.

Cavirani, Edoardo. In preparation. Epenthetic and intrusive vowels in Lunigiana: How diatopic variation mimics (phonologization) diachrony.

Cohn, Abigail & Sam Tilsen. 2015. Relation between syllable count judgments and durations of English liquid rimes. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Davidson, Lisa & Maureen Stone. 2003. Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. *Proceedings of the West Coast Conference on Formal Linguistics* 22. 165–178.

Dell, François & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7. 105–130.

Easterday, Shelece. 2017. *Highly complex syllable structure: A typological study of its phonological characteristics and diachronic development*. University of New Mexico. (Doctoral dissertation).

Fleischhacker, Heidi Anne. 2005. *Similarity in phonology: Evidence from reduplication and loan adaptation*. Los Angeles: University of California.

Gafos, Adamantios. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20(2). 269–337. DOI: 10.1023/A:1014942312445.

Garellek, Marc. 2020. Phonetics and phonology of schwa insertion in Central Yiddish. *Glossa: a journal of general linguistics* 5(1). 66. DOI: 10.5334/gjgl.1141.

Garmann, Nina Gram, Hanne Gram Simonsen, Pernille Hansen, Holm Elisabeth, POST Brechtje & Elinor Payne. 2021. Cross-linguistic variation in word-initial cluster production in adult and child language: Evidence from English and Norwegian. *Journal of Child Language* 48(1). 1–30.

Gick, Bryan & Ian Wilson. 2006. Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. *Laboratory phonology* 8. 635–659.

Grice, Martine, Michelina Savino & Timo B. Roettger. 2018. Word final schwa is driven by intonation: The case of Bari Italian. *The Journal of the Acoustical Society of America* 143(4). 2474–2486. DOI: 10.1121/1.5030923.

Hall, Nancy. 2003. *Gestures and segments: Vowel intrusion as overlap*. University of Massachusetts Amherst. (Doctoral dissertation).

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429. DOI: 10.1017/S0952675706000996.

Hall, Nancy. 2013. Acoustic differences between lexical and epenthetic vowels in Lebanese Arabic. *Journal of Phonetics* 41(2). 133–143.

Hall, Nancy, Andie Niederecker, Elica Sue & Irene Orellana. 2019. Annotating archival recordings of Hocank (Winnebago). In *Proceedings of the Annual Meetings on Phonology*, vol. 6.

Hamann, Silke & Veronica Miatto. 2024. Three language-specific phonological interpretations of release bursts and short vowel-like formants. In Ji Yea Kim, Veronica Miatto, Andrija Petrović & Lori Repetti (eds.), *Epenthesis and beyond: Recent approaches to insertion in phonology and its interfaces*, 199–223. Berlin: Language Science Press. DOI: 10.5281/zenodo.14264544.

Hammond, Michael, Natasha Warner, Andréa Davis, Andrew Carnie, Diana Archangeli & Muriel Fisher. 2014. Vowel insertion in Scottish Gaelic. *Phonology* 31(1). 123–153.

Harms, Robert T. 1976. The segmentalization of Finnish 'nonrules'. In *Texas linguistic forum*, vol. 5, 73–88.

Heselwood, Barry, Abdurraouf Shitaw, Aimen Ghummed & Leendert Plug. 2015. Epenthetic and excrescent vowels in stop sequences in Tripolitanian Libyan Arabic. In *Proceedings of the 18th International Congress of Phonetic Sciences.*

Hewson, John. 1986. Syllables and syllabics in Micmac. *Algonquian Papers-Archive* 17. 135–142.

Hickey, Raymond. 1985. The interrelationship of epenthesis and syncope: Evidence from Dutch and Irish. *Lingua* 65(3). 229–249.

Huang, Hui-chuan J. 2018. The nature of pretonic weak vowels in Squliq Atayal. *Oceanic Linguistics* 57(2). 265–288.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20(2). 219–273.

Karlin, Robin. 2021. Finnish inserted vowels: A case of phonologized excrescence. *Nordic Journal of Linguistics* 45(1). 49–79.

Kim, Kyumin & Alexei Kochetov. 2011. Phonology and phonetics of epenthetic vowels in loanwords: Experimental evidence from Korean. *Lingua* 121(3). 511–532.

Kitto, Catherine & Paul V. De Lacy. 1999. Correspondence and epenthetic quality. In C. Kitto & C. Smallwood (eds.), *Proceedings of AFLA (Austronesian Formal Linguistics Association VI*, 181–200. Toronto: University of Toronto, Dept. of Linguistics.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind, et al. 1998. Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28(1-2). 1–41.

Leonard, Janet. 2007. A preliminary account of stress in Senćoten (Saanich/North Straits Salish). *Northwest Journal of Linguistics* 1(4). 1–59.

Levin, Juliette. 1987. Between epenthetic and excrescent vowels. *Proceedings of the West Coast Conference on Formal Linguistics* 6. 187–201.

Miatto, Veronica. 2020. Perception of word-final inserted vowels and syllabicity in Italian. In *Proceedings of the 11th International Conference of Experimental Linguistics (ExLing 2020)*, 137–140.

Michelson, Karin. 1989. Invisibility: Vowels without a timing slot in Mohawk. In Donna Gerdts & Karin Michelson (eds.), *Theoretical perspectives on native American languages*, 38–69. Albany: Suny Press Albany.

Miner, Kenneth L. 1979. Dorsey's law in Winnebago-Chiwere and Winnebago accent. *International Journal of American Linguistics* 45(1). 25–33.

Ng, E-Ching. 2013. Paragoge as an indicator of language contact. *Extended Abstracts of the LSA Annual Meeting* 4(31). DOI: 10.3765/exabs.v0i0.794.

Nogita, Akitsugu & Yanan Fan. 2012. Not vowel epenthesis: Mandarin and Japanese ESL learners' production of English consonant clusters. *Working Papers of the Linguistics Circle* 22(1). 1–26.

Operstein, Natalie. 2010. *Consonant structure and prevocalization*. Amsterdam: John Benjamins.

Pariente, Itsik. 2010. Pharyngeal related non-lexical vowels in Sephardic Modern Hebrew. *Linguistics in Amsterdam* 3(1). https://linguisticsinamsterdam.nl/home?issue=142.

Repetti, Lori. 2012. Consonant-final loanwords and epenthetic vowels in Italian. *Catalan Journal of Linguistics* 11. 167–188.

Ridouane, Rachid & Jamison Cooper-Leavitt. 2019. A story of two schwas: A production study from Tashlhiyt. *Phonology* 36(3). 433–456.

Ring, Hiram & James Gruber. 2014. Intrusive vowels in Pnar: There and not there. In *24th meeting of the SouthEast Asian Linguistics Society (SEALS 24)*. Yangon University.

Smith, Kaylin. 2019. Vowel insertion in Scottish English liquid + nasal clusters: Boundary disputes and durational relations. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Vowel insertion in Scottish English liquid + nasal clusters: Boundary disputes and durational relations*, 132–136. Melbourne: International Phonetic Association.

Steriade, Donca et al. 1990. Gestures and autosegments: Comments on Browman and Goldstein's paper. *Between the grammar and the physics of speech: Papers in laboratory phonology* 1. 382–397.

Susman, Amelia. 1943. *The accentual system of Winnebago*. Columbia University. (Doctoral dissertation).

Tabain, Marija & Birgit Hellwig. 2022. Qaqet. *Journal of the International Phonetic Association* 53(3). 1073–1094.

Van Donselaar, Wilma, Cecile Kuijpers & Anne Cutler. 1999. Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of memory and language* 41(1). 59–77.

Walter, Mary Ann. 2018. The significance of vowel insertion in Ottoman Turkish. In *The Rouen Meeting: Studies on Turkic Structures and Language Contacts*, 1st edn., 87–94. Harrassowitz Verlag.

Warner, Natasha, Allard Jongman, Anne Cutler & Doris Mücke. 2001. The phonological status of Dutch epenthetic schwa. *Phonology* 18(3). 387–420.

Wiik, Kalevi. 1965. *Finnish and English vowels: A comparison with special reference to the learning problems met by native speakers of Finnish learning English*, vol. 94. Turun: Turun yliopisto.