Phonology **36** (2019) 1–36.
doi:10.1017/S0952675719000022

*Modelling incomplete neutralisation with weighted phonetic constraints**

Aaron Braver
Texas Tech University

Incomplete neutralisation presents a problem for classical modular feed-forward grammars: it results in surface phonetic distinctions between phonologically neutralised segments. This paper argues for a model of incomplete neutralisation using two independently motivated theoretical devices: paradigm uniformity and weighted phonetic constraints. A case study is presented, showing that Japanese monomoraic lengthening results in incomplete neutralisation: when monomoraic nouns with short vowels are lengthened to fill a bimoraic minimality requirement, they reach a duration intermediate between that of unlengthened short vowels and underlyingly long vowels. The Japanese case has properties distinct from other classically cited examples of incomplete neutralisation such as final devoicing, which are not predicted by previous theories of neutralisation. The Weighted Paradigm Uniformity theory of incomplete neutralisation is shown to make four unique predictions, and is argued to better capture the typology of incomplete neutralisation.

1 Introduction

Since at least the early 1980s, linguists have noted that *phonological* neutralisation does not always yield *phonetically identical* segments (Mitleb 1981b, Port *et al*. 1981, Port & O'Dell 1985). In this 'incomplete' neutralisation, phonological neutralisation is complete (i.e. the segments are assumed to share an identical phonological representation), yet a phonetic trace of the underlying phonological distinction is present in the surface form. Incomplete neutralisation presents a challenge to traditional modular feed-forward models (Chomsky & Halle 1968, Keating 1996, Pierrehumbert 2002, Bermúdez-Otero 2007), in which only the output of phonology can influence phonetic realisation.[1] In incomplete

* E-mail: AARON.BRAVER@TTU.EDU.
I thank Shigeto Kawahara and Will Bennett for their discussion of this project. I am also grateful to three anonymous reviewers and an associate editor for providing substantive comments and direction that greatly improved the paper. The Japanese data presented in §3 come from Braver & Kawahara (2014, 2016).

[1] Modular feed-forward models are grammatical architectures in which the lexicon, phonology and phonetics occupy separate modules, and in which information can

neutralisation, it appears that the underlying contrast between forms – which is supposedly obscured from the phonetic module after phonological neutralisation occurs – still plays a role in phonetic realisation. While the classical model predicts that phonologically neutralised sounds should be realised identically, incomplete neutralisation yields surface distinctions between phonologically neutralised segments.

In spite of the predictions of the traditional model, incomplete neutralisation has been found in a wide variety of languages and phenomena. Perhaps the most frequently cited cases of incomplete neutralisation involve processes of final devoicing – a voicing contrast is phonologically neutralised, but phonetic distinctions are evident between devoiced segments and their underlyingly voiceless counterparts (see §1.1 for further details and an extensive list of cases of incomplete neutralisation in both processes of devoicing and in other contexts).

Due to this apparent contradiction, incomplete neutralisation has been called a 'theoretical bad dream' (Manaster Ramer 1996a, Port 1996). In this paper, I argue that incomplete neutralisation is not an intractable problem, but rather that it can be viewed as a tension between two independently motivated grammatical forces: faithfulness to a base form (Benua 1997, Steriade 2000, Albright 2002b) competes with adherence to a segment's canonical realisation. This tension is modelled in a phonetic grammar which uses weighted phonetic constraints (Legendre *et al*. 1990, Zsiga 2000, Flemming 2001), which I term the WEIGHTED PARADIGM UNIFORMITY theory of incomplete neutralisation.

The remainder of the paper is organised as follows. §1.1 defines and describes several cases of incomplete neutralisation. In §2, I present Japanese monomoraic noun lengthening as a case study of incomplete neutralisation. §3 lays out existing theories of incomplete neutralisation, and provides theoretical background for the model proposed in this paper, which is described in §4. In §5 I discuss the predictions of this model as compared to other theories of incomplete neutralisation.

1.1 Incomplete neutralisation

The best-studied case of incomplete neutralisation is undoubtedly final devoicing in German (Mitleb 1981a, b, Port *et al*. 1981, Port & O'Dell 1985, Röttger *et al*. 2014, as well as Taylor 1975 (for some places of articulation) and Dinnsen & Garcia-Zamor 1971 (for disyllables only)). Under the assumptions of the modular feed-forward model, German /ʁat/ 'advice' and /ʁad/ 'wheel' should surface identically, given that the final obstruents are both voiceless in the phonological output. Contrary to this expectation, a trace of the underlying voicing distinction is present

pass from the lexicon to phonology and from the phonology to the phonetics, but not *vice versa*: 'no arrows go backwards, from articulatory plans to phonological encoding or from phonological encoding to the lexical level' (Pierrehumbert 2001b: 102).

on the surface: the vowel in /ʁad/ is longer than the one in /ʁat/.[2] More specifically, Port & O'Dell's (1985) study of German final devoicing found that vowels preceding devoiced final obstruents were approximately 15 ms longer than vowels preceding underlyingly voiceless obstruents (among other surface distinctions).

With this example in hand, we can now provide the definitions of both complete and incomplete neutralisation in (1).[3]

(1) a. *Complete neutralisation*
The surface acoustic realisation of the contrast between two underlyingly distinct segments (in a given context) is completely identical.

b. *Incomplete neutralisation*
The surface acoustic realisation of the contrast between two underlyingly distinct segments (in a given context) is less distinct than the segments' canonical realisations in a non-neutralising context (i.e. some degree of neutralisation has occurred), but they are not completely identical.

If, contrary to fact, German final devoicing were completely neutralising, we would expect that the surface acoustic realisation of the voicing contrast between /ʁad/ and /ʁat/ (e.g. duration of preceding vowels) should be identical. As the facts actually are – that this contrast is incompletely neutralised – the surface acoustic cues to the voicing contrast are distinct, yet not as distinct as the realisation of the voicing contrast in non-neutralising contexts. As noted above, the difference in preceding vowel duration between devoiced (underlyingly voiced) and underlyingly voiceless final obstruents is only about 15 ms, as compared to

[2] Preceding vowel duration is one of the many cues to a voicing distinction: cross-linguistically, vowels preceding voiced segments are longer than those preceding voiceless segments. Other cues to voicing include closure duration (Kluender *et al*. 1988), effects on the F0 and F1 of surrounding vowels (Hombert *et al*. 1979, Kingston & Diehl 1994) and F1-F2 divergence in the offglides of closing diphthongs (Thomas 2000, Moreton 2004). The majority of work on incomplete neutralisation in final devoicing has focused on vowel duration, hence its use in the examples in this section.

[3] In the definitions of complete and incomplete neutralisation, I use the term 'segment' for the sake of simplicity. As in the case of the near-merger of Cantonese tone (Yu 2007) and monomoraic lengthening in Japanese (on which see §2), the unit involved may be suprasegmental.

I assume, also for the sake of simplicity, that there is a Distinctness function which takes as arguments two phonological units and the feature on which to compare them, and returns a value where 0 is identity and which has increasing values corresponding to increasing distinctness. I leave open whether the Distinctness function measures raw acoustic properties or taps into speakers' knowledge of contrasts, in the manner of the P-map (Steriade 2009).

In incomplete neutralisation, the reduction of contrast is not due purely to the physical mechanics of articulation. As pointed out by an anonymous reviewer, the VOT contrast is decreased before low vowels as compared to high vowels; the link between VOT and vowel height is likely mechanical (Chang *et al*. 1999, Koenig *et al*. 2011), rather than an effect of a speaker's grammar *per se*. Such cases are not normally considered to be incomplete neutralisation.

non-neutralising contexts in which this duration difference may be more than doubled (Mitleb 1981a, b).[4]

Incomplete neutralisation of final devoicing has also been reported in Catalan (Dinnsen & Charles-Luce 1984), Dutch (Warner *et al.* 2004, though see the article itself and Warner *et al.* 2006 for caveats), Russian (Dmitrieva 2005, Kharlamov 2012, Matsui 2015) and Polish (Gianinnia & Cinque 1978, Slowiaczek & Dinnsen 1985, Slowiaczek & Szymanska 1989, though see Jassem & Richter 1989 for counterevidence). In addition to incomplete neutralisation in final devoicing, incomplete neutralisation has been found in flapping in American English (Fisher & Hirsh 1976, Fox & Terbeek 1977, Zue & Laferriere 1979, Herd *et al.* 2010, Braver 2014), morphological tone in Cantonese (Yu 2007),[5] coda aspiration in Eastern Andalusian Spanish (Gerfen 2002) and monomoraic noun lengthening in Japanese (Braver 2013, Braver & Kawahara 2014, 2016).

The case of monomoraic noun lengthening in Japanese will be discussed in detail in §2, though at this point it is worth mentioning two features of this process that differ from final devoicing. First, Japanese monomoraic noun lengthening is a more convincing case of truly phonological neutralisation than final devoicing. In the cases of final devoicing cited above, there is not clear evidence to confirm that 'devoiced' segments are treated as voiceless by independent phenomena (indeed, Barnes 2006 argues that incomplete neutralisation of final devoicing should be treated as a purely phonetic affair – that at the phonological level, devoiced and underlyingly voiceless segments remain distinct, but happen to have phonetic targets which are nearly identical). In the Japanese case, as will be shown in more detail in §2, lengthened monomoraic nouns are treated the same as underlyingly bimoraic nouns for purposes of a bimoraic minimality constraint, and can carry pitch accent – a property reserved for words of two moras or greater.

Second, there are a large number of phonetic correlates to voicing, including duration of the preceding vowel, the correlate most frequently discussed with respect to incomplete neutralisation in final devoicing (Chen 1970), closure duration (Kluender *et al.* 1988) and F0 and F1 (Kingston & Diehl 1994, Hombert *et al.* 1979). In the Japanese case, however, the primary correlate of vowel length is (unsurprisingly) vowel duration (see e.g. Hirata 2004), with only secondary non-durational

4 Port & O'Dell (1985) also found differences between devoiced and underlyingly voiceless segments in aspiration duration and voicing duration in consonant closure, as well as a marginal difference in closure duration.

5 Incomplete neutralisation of morphological tone in Cantonese is reported by Yu (2007) as a case of 'near-merger'. The distinction between 'incomplete neutralisation' on the one hand and 'near-merger' on the other is not clear-cut. Historically, the term 'incomplete neutralisation' has been more commonly used in the phonetic and phonological literature, while 'near-merger' is found more often in the sociolinguistic literature – as such, near-merger tends to be used to describe cases of subphonemic distinctions that have resulted from recent sound changes, whereas incomplete neutralisation describes a situation with synchronic phonological alternations.

correlates (Behne *et al.* 1999, Kinoshita *et al.* 2002, Hirata & Tsukada 2009), thus simplifying the analysis.

2 Japanese monomoraic noun lengthening

In this section, I describe the phenomenon of incomplete neutralisation in Japanese monomoraic noun lengthening (Mori 2002, Braver & Kawahara 2014, 2016). This case will be used in §4 to illustrate the proposed model.

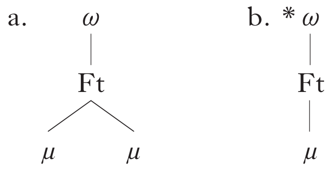
2.1 Background: bimoraicity requirement

Japanese requires that all prosodic words (ω) have at least two moras (Itô 1990, Mester 1990, Poser 1990, Mori 2002, Itô & Mester 2003). This bimoraicity requirement is observed in many word-formation patterns, all of which are based on a bimoraic template, including nickname formation, geisha-client name formation, loanword abbreviation, verbal root reduplication, scheduling compounds and telephone number recitation.

For instance, in the nickname formation pattern, a full name must be truncated to two moras before the suffix *-chan* can be applied.[6] For example, the five-mora name *Wasaburoo* can be truncated to two moras, as in (2a), but not one. Similarly, the three-mora name *Kotomi* can be truncated to either two monomoraic syllables, or a single bimoraic syllable, as in (2b). It cannot, however, be shortened to a single mora.

(2) a.	Wasaburoo	(full name)	b.	Kotomi	(full name)
	Wasa(-chan)	(2 moras)		Koto(-chan)	(2 moras)
	*Wa(-chan)	(1 mora)		Koc(-chan)	(2 moras)
				*Ko(-chan)	(1 mora)

To summarise, a prosodic word must contain at least one foot, and the foot must be binary (at the moraic level in Japanese; McCarthy & Prince 1986, 1993), as in (3).

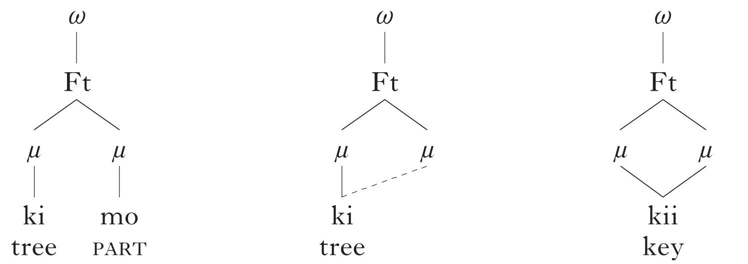
(3)


In spite of this bimoraicity requirement, there are monomoraic nouns in the Japanese lexicon; e.g. [ki] 'tree', [i] 'stomach' and [e] 'picture'. When these monomoraic nouns appear in isolation within a prosodic word (e.g. without a case particle), they are produced with vowels of intermediate length – longer than canonical short vowels, but shorter than underlyingly

[6] Here and throughout, Japanese morphemes are given in the standard Romaji romanisation.

long vowels (Mori 2002, Braver & Kawahara 2014, 2016). Underlyingly bimoraic nouns in the same environment do not show such lengthening. Mori thus concludes that the lengthening is caused by the phonological bimoraic minimality requirement: monomoraic nouns with a case particle within their ω satisfy the bimoraicity requirement (by virtue of the particle's mora), as in (4a), while monomoraic nouns without a particle must gain an additional mora to satisfy this requirement, as in (4b). No lengthening is required for underlyingly bimoraic nouns, as in (4c).

(4) a. *No lengthening with a particle* b. *Lengthening without a particle* c. *Underlyingly bimoraic nouns*



2.2 Monomoraic lengthening is incompletely neutralising

In spite of the identical surface mora counts in (4b) and (4c) (due to lengthening in the case of (4b)), Braver & Kawahara (2016) found that the vowel durations of lengthened monomoraic nouns were not identical to those of underlyingly long nouns. In that study, twelve native speakers of Japanese were presented with 15 sets of three nouns: (a) a short noun with a case particle (no lengthening), (b) a short noun without a case particle (lengthening expected) and (c) an underlyingly long noun. (5) shows a sample triplet, with nouns in a frame sentence. The target nouns (the first word in each frame sentence) for all three members of the triplet share identical segmental material, differing only in the presence/absence of a case particle and underlying vowel length.

(5) *Sample stimulus set* (from Braver & Kawahara 2016)

	condition	*orthography*	
a.	short, with particle	木もなくしたよ。	ki mo nakushita yo tree also lost DISC
b.	short, no particle	木なくしたよ。	ki nakushita yo tree lost DISC
c.	long	キーなくしたよ。	kii nakushita yo key lost DISC

Braver & Kawahara (2016) found that lengthened nouns were on average 32.47 ms shorter than underlyingly long nouns, as summarised in Table I. In other words, the short/long vowel contrast is incompletely neutralised:

lengthened short vowels are more like long vowels than short vowels in non-neutralising contexts, but they are not identical to long vowels.

condition	mean	SD	rounded
unlengthened short (with particle)	54.99	21.89	50
lengthened short (without particle)	124.98	34.91	125
underlyingly long (without particle)	157.45	39.21	150

Table I

Means, standard deviations and rounded values for vowel duration of nouns (in ms), from Braver & Kawahara (2016) (12 speakers, 15 sets of 3 nouns (= 45 items in total), 7 repetitions).

For ease of explication, in §4 I will focus on the overall means (across speakers and vowel qualities), and will use the rounded values for the target durations of short, lengthened and long vowels given in the right-most column of Table I.

2.2.1 *Bimoraic nouns with and without case particles.* It should be noted that the nouns used in the underlyingly long condition in the Braver & Kawahara (2016) study were not followed by a case particle (e.g. *kii*, rather than *kii-mo*). An anonymous reviewer points out that in order for the length contrast to be incompletely neutralised, the duration difference between monomoraic nouns with and without a particle (e.g. *ki-mo* ~ *ki*, involving degree of lengthening) must be smaller than the duration difference between monomoraic and bimoraic nouns with particles (e.g. *ki-mo* ~ *kii-mo*, with an underlying length contrast). In other words, to be an actual case of incomplete neutralisation, the degree of lengthening in putatively incompletely neutralised lengthened monomoraic nouns must be less than the normal length contrast found in the non-neutralising context of monomoraic *vs*. bimoraic nouns with particles.

There are two reasons to believe that the degree of lengthening (69.99 ms; see Table I) is indeed less than the underlying length contrast. First, Mori (2002) shows that bimoraic nouns hardly differ in duration when uttered with *vs*. without particles – those without case particles are only longer by 5–6 ms (4–5%). Even if we assume that particle-less *kii* is 10% longer than *kii-mo*, we should expect *kii-mo* to be approximately 143 ms. Assuming this generous estimate, the underlying length contrast is 88.15 ms – a great deal larger than the degree of lengthening in monomoraic nouns.

Second, an earlier version of the Braver & Kawahara (2016) study included a condition with bimoraic nouns with particles (e.g. *kii-mo*; Braver & Kawahara 2014). In this first experiment, such bimoraic nouns were on average 26.55 ms longer than lengthened particle-less monomoraic nouns (e.g. *ki-mo* ~ *ki*) – much smaller than the underlying contrast between monomoraic nouns with particles and bimoraic nouns with particles (e.g. *ki-mo* ~ *kii-mo*) of 72.20 ms. These results are summarised in Table II.

condition	mean	SD
unlengthened short (with particle)	73.54	20.58
lengthened short (without particle)	119.19	32.56
underlyingly long (without particle)	145.74	31.21

Table II

Means and standard deviations for vowel duration of nouns (in ms), from Braver & Kawahara (2014) (7 speakers, 11 sets of 3 nouns (= 33 items in total), 10 repetitions).

2.2.2 *Lengthened nouns are phonologically bimoraic.* As discussed in the introduction, monomoraic noun lengthening is perhaps a more compelling case of truly phonological neutralisation than final devoicing. There are two pieces of evidence to support this claim: (i) lengthened monomoraic nouns are allowed in phonological contexts which require bimoraic minimality, and (ii) lengthened monomoraic nouns can carry a pitch accent.

Japanese pitch accent is realised as H*L – a high tone on the accented mora, followed by a low tone realised on the following mora. Since the tone-bearing unit in Japanese is the mora (Haraguchi 1977, McCawley 1977), any word which manifests both the H* and L tones of the Japanese pitch accent must be minimally bimoraic. This property is exemplified in the pitch tracks in Fig. 1 for the sentence *sho'/sho'u* (*mo*) *dokusenshita yo* 'book/chapter (also) monopolised DISC'. (Following Japanese transliteration conventions, accent is marked as <'> between the moras expressing H and L tones, and the long vowel in *sho'u* is represented as *ou*, but pronounced [oː].)

In the pitch track for *sho'u dokusenshita yo* in Fig. 1a, the start of the long vowel [oː] is realised as a high tone, descending to a low tone by the end of the second mora. Similarly, in *sho' mo dokusenshita yo* in Fig. 1b, the high tone is realised on the first (and only) mora of *sho*, while the descent to a low tone extends beyond the word boundary and into the mora of *mo*. If lengthened monomoraic nouns can show pitch accent (and are therefore phonologically two moras), we should expect the fall from H* to L in *sho' dokusenshita* to spread across the two moras of the lengthened *o* – and this is precisely what occurs (Fig. 1c).

In addition to the pitch-accent data, there is additional evidence that the particle-less context described above does indeed require two moras. If a

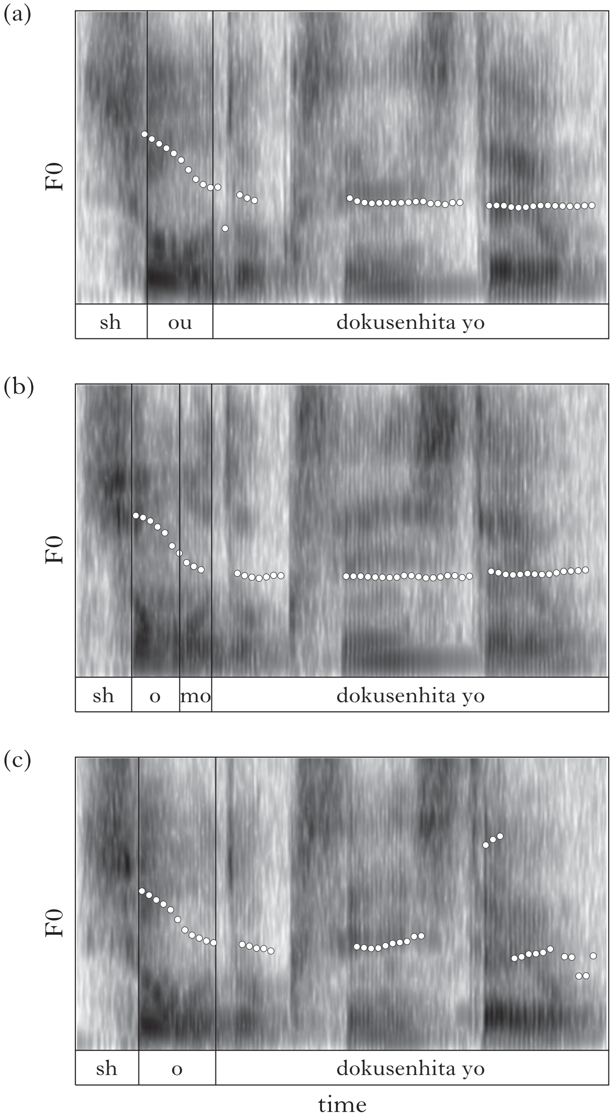


Figure 1

Pitch tracks for Japanese nouns in a carrier sentence (male speaker reading *sho'*/*sho'u* (*mo*) *dokusenshita yo* 'book/chapter (also) monopolised DISC'): (a) underlyingly long; (b) short with particle; (c) lengthened short noun (no particle).

monomoraic noun without a particle is followed by a glottal stop or a pause, these additions can provide the relevant ω with the requisite mora (Mori 2002). Mori argues that this possibility is allowed because the additional glottal stop or pause takes up a similar amount of time as a second mora would.

3 Theories of incomplete neutralisation and background

3.1 Previous accounts

Perhaps the earliest theoretical account of incomplete neutralisation is due to Anderson (1975), who argues that some phonetic rules can precede phonological ones. Without using the term incomplete neutralisation, Anderson points out that flapping in American English, where /t/ and /d/ become /ɾ/ in certain prosodic contexts, leads to a gradient distinction in the duration of the preceding vowel: vowels preceding /d/-flaps are slightly longer than vowels preceding /t/-flaps. He further notes that a standard rule-based analysis of this phenomenon, as in (6), cannot account for the data.

(6) a. V → [+long] / __ [+voice]   b. {t,d} → [ɾ] / V __ V̆

Under the analysis in (6), vowels lengthen first, followed by /t/ and /d/ becoming flaps. This ordering – which is necessary to ensure that vowels are lengthened only before /d/-flaps and not /t/-flaps – contradicts the modular feed-forward model, since the gradient (and hence phonetic) vowel-lengthening rule precedes the (phonological) flapping rule. In order to resolve this conflict, Anderson proposes that (some) phonetic rules may, in fact, precede (some) phonological rules.

More recently, van Oostendorp (2008) has argued that incomplete neutralisation in German final devoicing can be accounted for by ensuring that devoiced (underlyingly voiced) and underlyingly voiceless segments have distinct phonological representations. Under this analysis, based on Turbidity Theory (Goldrick 2000), segments can stand in two possible relations with a given feature: (i) the PROJECTION relationship, which is an 'abstract, structural relationship', and (ii) the PRONUNCIATION relationship, which 'describes the output realisation of structure' (van Oostendorp 2008: 1368). The turbidity model analyses incomplete neutralisation in coda devoicing by providing different structures for underlyingly voiceless and voiced segments: in the phonological output, devoiced segments maintain a projection relation (but not a pronunciation relation) with a [voice] feature, but underlyingly voiceless segments are not related by projection or pronunciation to [voice]. Since underlyingly voiceless and devoiced segments are thus distinct in the phonological output, the phonetic module can differentiate between them, allowing longer vowels to precede devoiced segments, for example.

Another proposal, due to Gouskova & Hall (2009), capitalises on the idea of a segment's derivational history. They present a study showing that epenthetic vowels in Lebanese Arabic are either shorter, backer or both shorter and backer than their lexical counterparts. In order to model this phenomenon in a phonological grammar, they argue that the phonetics must have access to an intermediate stage of phonological derivations. Following the assumptions of Optimality Theory with Candidate Chains (OT-CC; McCarthy 2007: ch. 3), they take a candidate to consist of a derivational chain from the phonological input to the phonological surface form, with gradual, incremental steps along the way. Under this approach, the epenthetic vowel [i] goes through a number of steps – starting as null, then gradually becoming more and more like a full vowel, forming a chain like the one in (7) (Gouskova & Hall's (8)).

(7) *Candidate chain for epenthesis of* [i]
   /CC/ <CC, CɨC, CəC, CiC>

If the phonetics can access the entire chain, they argue, rather than just the last step in the chain, epenthetic vowels in Lebanese Arabic might be realised as one of the other steps in the chain – explaining why some speakers produce the epenthetic vowel more like [ɨ] or [ə], rather than [i]. Underlying /i/, however, has no such chain of changes, and therefore must surface as canonical [i].

A different class of analysis, which I draw on in the model presented below, concerns paradigm uniformity among morphologically related forms (e.g. Benua 1997, Steriade 2000, Yu 2007). The effects of paradigm uniformity can be observed when a given allomorph avoids a language-general pattern in favour of similarity to some morphological neighbour. Steriade (2000) argues, for example, that words within a morphological paradigm share certain phonological and phonetic properties. She describes the (optional) process of schwa deletion in French, which renders forms such as (8a) (with surface schwa) as (8b) (schwa deleted).

(8) a.	*bas retrouvé*	[ba <u>ʁ</u>ətʁuve]	'stocking found again'
b.	*bas r'trouvé*	[ba <u>ʁ</u>tʁuve]	'stocking found again'
c.	*bar trouvé*	[ba<u>ʁ</u> tʁuve]	'bar found'

Crucially, the underlined [ʁ] in (8b) (with schwa deletion) is not phonetically identical to the one in (8c), where there was no underlying schwa. The [ʁ] in (8b) surfaces with 'qualities that would only be appropriate if the schwa was still present' (Steriade 2000: 327; cf. Rialland 1986, Fougeron & Steriade 1997). On Steriade's phonetic analogy analysis, forms that have undergone schwa deletion as in (8b) are influenced by forms like (8a), with surface schwa, thus accounting for their similarity.

An assumption of many theories of incomplete neutralisation is that speakers have relatively fine-grained control over phonetic implementation. In the model proposed by Yu (2011), the phonology has just such control over phonetic implementation of contrast. Following Kingston & Diehl (1994: 420, n.2), an allophone is simply 'any phonetic variant of a distinctive feature specification or arrangement of such specifications that occurs in a particular context'. That is to say, contrastive features can vary considerably in their realisation, depending on their context. For example, [+voice] in English may be realised with closure voicing intervocalically, but as voiceless unaspirated word-initially. Kingston & Diehl (1994) argue that speakers can therefore choose between the various methods to articulate a feature such as [+voice]. Yu (2011) claims that subphonemic differences, such as incomplete neutralisation and near-merger, should be seen in a similar light. In incomplete neutralisation of final devoicing, the non-final voiced 'allophone' and the word-final voiceless 'allophone' are distinct, even if their phonetic cues are so impoverished as to 'escape detection by traditional methods of linguistic data collection … Nonetheless, the contrast is maintained from the perspective of the native speaker, albeit covertly' (Yu 2011: 311).

A final model of incomplete neutralisation makes use of 'phonetic traces'. In their study of tongue-twister errors, Goldrick & Blumstein (2006) show that 'traces' of an intended target can have an effect on pronunciation. For example, in the mistaken use of [g] rather than [k], VOT is longer than in intended pronunciations of [g] – the mistaken production of [g] shows a trace of the intended [k] target. This phenomenon is modelled in a system with 'cascading activation' – information flows unidirectionally from the phonological module to the phonetic module; however, earlier stages in this process can generate and transmit multiple representations downstream. In the case of mistaken [g] for [k], representations of both [g] and [k] are activated at the phonological level and transmitted to the phonetics. Goldrick & Blumstein (2006) argue that this sort of analysis can be applied to incomplete neutralisation: in German final devoicing, a devoiced [d̥] is pronounced in a 't-like' fashion, due to partial activation of the underlying /t/ being sent to the phonetic module.

3.2 Weighted phonetic constraints and paradigm uniformity

The model presented in this paper relies on two independently motivated theoretical mechanisms: weighted phonetic constraints (Legendre *et al.* 1990, Zsiga 2000, Flemming 2001) and paradigm uniformity (Benua 1997, Steriade 2000). These are discussed in the following sections.

3.2.1 *Weighted phonetic constraints.* A crucial component of the model is the idea that competing phonetic demands must reach a compromise. This idea is familiar from the phonetic literature: both Zsiga (2000) and Flemming (2001) argue for weighted constraint grammars (Legendre *et al.* 1990, Pater 2009) that can operate over phonetic information. Zsiga (2000) argues for a set of 'phonetic alignment constraints', which apply separately from the phonology; on the other hand, Flemming (2001) articulates a unified model of phonetics and phonology, deriving both categorical and gradient phenomena from the same set of weighted constraints.

Key to both these systems is the use of weighted constraints, as opposed to strictly dominating ranked constraints as in classical Optimality Theory (Prince & Smolensky 1993). Crucially, weighted constraints can generate compromise. With ranked constraints, conflict is resolved by acceding to the demands of the highest-ranked constraint. Under weighted constraints, neither conflicting constraint must be satisfied categorically in order to yield an optimal output. The candidate which represents the best compromise between two highly weighted constraints may yield a lower overall cost (or equivalently, higher overall harmony) than a candidate which follows completely the demands of any one constraint. As argued by Zsiga (2000: 96–97), this attribute of weighted constraint systems makes them optimal for the analysis of gradient phenomena.

The two models differ, however, in their relationship between phonological and phonetic processes. Zsiga (2000) envisions a model with two constraint-based grammars: a phonological grammar first manipulates phonological representations (in terms of, for example, features, segments and prosodic units), followed by a phonetic component which takes the output of the first grammar as its input, and assesses candidates which consist of phonetic realisations in terms of gestural targets. Constraints in the phonetic component can require alignment of gestural targets, or specify goals such as 'be distinct' or 'conserve energy'.

Flemming's (2001) proposal is conceptually similar to Zsiga's model, with the major distinction that the phonetic and phonological components are merged. Constraints in this joint phonetics/phonology module can refer both to phonological structures and to raw phonetic values (in milliseconds, Hertz, etc.). For example, in order to model cross-linguistic differences in coarticulatory effects, competing constraints pressure segments to match phonetic targets (e.g. a particular F2 value) on the one hand, and to be similar to neighbouring segments on the other. Varying the weight of these constraints determines how the conflict is resolved.[7]

For the sake of concreteness, I assume a two-grammar model along the lines of Zsiga (2000), in which an OT-like grammar implements phonological processes first, followed by a weighted-constraint phonetic grammar which takes the output of phonology as its input. Further, I assume that phonetic constraints may make reference to both phonological categories and raw phonetic details – in particular, raw phonetic duration. As such, the constraints defined below are similar in spirit to Flemming's (2001) $F2(C)=L$ and $F2(V)=T$ constraints, in which a segment is gradiently penalised for phonetic distance from its target: for example, a consonant that misses its F2 target by 200 Hz is assigned a cost of 200 by $F2(C)=L$, and a vowel that misses its F2 target by 50 Hz is assigned a cost of 50 by $F2(V)=T$.

[7] An anonymous reviewer points out that Flemming (2002) argues for an analysis of final devoicing in terms of categorically assessed MINDIST constraints. I follow the approach in Flemming (2001), and assume gradiently assessed constraints.

3.2.2 *Paradigm uniformity*. Paradigm uniformity, as discussed in §3.1, requires morphologically related forms to be faithful to one another phonologically (Benua 1997) and/or phonetically (Steriade 2000). I formalise the phonetic pressure to remain faithful to a base in the constraint OO-ID(dur), defined below in (18).

The effects of paradigm uniformity are evident when a given form avoids a language-general pattern in favour of similarity to a morphological relative. For example, in Canadian Raising, vowels preceding voiceless segments generally raise from /aɪ/ to [ʌɪ] (e.g. *ride* [ɹaɪd] ~ *write* [ɹʌɪt]). This raising overapplies in the case of *writer* /ɹaɪtɹ̩/ → [ɹʌɪɾɹ̩], *[ɹaɪɾɹ̩], even though the vowel precedes a voiced segment (a flap). Under a paradigm-uniformity analysis, the language-general pattern in which /aɪ/ does not raise before voiced segments is violated so that the vowel in *writer* can match the vowel in its morphologically related base, *write* [ɹʌɪt].

Under Benua's (1997) classic formulation, a candidate must be faithful to its 'base' – a form that is in the same paradigm as the candidate and is less morphologically complex than the candidate (e.g. *penguin* can serve as the base for *penguin-like*, but not *vice versa*). In §3.3 I discuss this model of basehood, and compare it to alternative proposals, arguing that informativeness, rather than morphological complexity, is the most appropriate method of base selection for Japanese monomoraic lengthening.

3.3 Choosing a morphologically related base

There are four major approaches to choosing the base to which a form must be faithful: morphological complexity, orthography, frequency and informativeness. In the Japanese case, the base to which a monomoraic noun must be faithful is the form with a particle (e.g. *ki mo*), not the bare form (e.g. *ki*) – this ensures that when monomoraic nouns lengthen, they are prevented from becoming as long as underlyingly long nouns. This section briefly describes these approaches, and argues that informativeness most unambiguously selects the form with particles as the base in monomoraic lengthening.

3.3.1 *Morphological complexity and orthography*. As noted above, Benua (1997) argues that basehood is determined by relative degree of morphological simplicity: a base must be a licit word which is morphologically less complex than the target word. Depending on one's analysis of the structure of particle-less nouns, basehood based on morphological simplicity may be incompatible with the facts of Japanese monomoraic lengthening. Assuming that nouns without particles have no additional structure (e.g. they don't have a particle which just happens to be silent), the particle-less noun is morphologically simpler than a noun with a particle. In this case, $[\text{ki Ø}]_\omega$ would either have to serve as its own base, or have no base at all.

A second approach to basehood, especially in the literature on incomplete neutralisation, is an appeal to orthography. It has been argued that incomplete neutralisation is the result of interference from speakers' orthographic knowledge or 'hypercorrect spelling pronunciation' (Fourakis & Iverson 1984: 142; see also Manaster Ramer 1996b, Piroth & Janker 2004, Warner *et al*. 2006). As an example, consider again English *writer* and *rider*: the medial stop neutralises to [ɾ], but traces of the underlying word are detectable on the surface (i.e. vowels before /d/-flaps are longer than those before /t/-flaps). The orthography-based approach argues that this surface difference is caused not by reference to the underlying voicing status of the medial consonant, nor by influence from paradigmatically related forms like *write* and *ride*, but rather by the speaker either seeing the orthographic *t* or *d* at the time of utterance, or by recalling the orthographic contrast at the time of utterance. In this way, the orthographic form serves as the base.

One argument against this orthographic approach in general is that some languages, such as Catalan (Dinnsen & Charles-Luce 1984), do not orthographically encode the contrast of interest in relevant positions, yet still exhibit incomplete neutralisation of these contrasts (see Kharlamov 2012 for further discussion). In the case of Japanese monomoraic lengthening, too, there are reasons to think that orthography is not the main factor behind incomplete neutralisation, and that orthographic forms do not serve as bases. In the Braver & Kawahara (2016) study, out of 13 short/long noun pairs, only one short noun was written in a phonetic script, the katakana syllabary: ソ *so* '5th note of the diatonic scale'. In this instance, it is plausible to argue that the base of a lengthened *so* is its orthographic representation ソ, and, as such, the degree of lengthening might be diminished, in allegiance to the short syllable represented in the orthography. In most cases, though, target nouns were presented in kanji – a primarily logographic system which does not directly encode phonological length (e.g. 麩 *fu* 'gluten' ~ 封 *fuu* 'seal'). In the case of lengthened *fu*, the orthographic base would be 麩, which does not encode length, and as such should not influence duration.

3.3.2 *Frequency*. Another approach to basehood relies on frequency: the most frequent member of a morphological paradigm serves as the base (Mańczak 1958). To examine whether this approach is appropriate in the case of Japanese monomoraic noun lengthening, the NINJAL Parsed Corpus of Modern Japanese (NPCMJ) was searched for nouns and their immediately following morpheme.[8] Of the 6550 nouns in the corpus, 5786 (88.34%) were immediately followed by a particle, while the remaining 764 nouns were followed by some other part of speech, such as another adjective, a noun or number (i.e. not a particle). The five most frequent particles are shown in Table III.

[8] The NPCMJ, available at http://NPCMJ.ninjal.ac.jp/, contains data from articles in the Kahoku Shimpo newspaper, selected Wikipedia articles and two translations of the bible into Japanese. The bible translations were excluded, so as to avoid highly formal and sometimes archaic language.

o (ACC)	*no* (GEN)	*ni* (DAT)	*ga* (NOM)	*wa* (TOP)	Ø
1121 17.11%	1002 15.30%	965 14.73%	847 12.93%	672 10.26%	764 11.66%

Table III
Frequency of most common particles in the NPCMJ corpus. Percentages indicate frequency among all nouns ($n=6550$).

Given that the caseless nouns are less frequent than those followed by *o*, *no*, *ni* or *ga*, the frequency approach to basehood predicts, given the data above, that the base for Japanese nouns should be a noun with a case particle – the desired outcome. Putting aside for the moment which particular case particle attaches to the base, the base of a monomoraic noun will always contain sufficient material to meet the bimoraicity requirement, by virtue of the case particle.

One question raised by this analysis, pointed out by an anonymous reviewer, is whether base frequency is computed as the cumulative frequency of all suffixed forms *vs*. unsuffixed forms, or whether one particular suffixed form must be more frequent than unsuffixed forms. In the analysis presented above, any one of *o*, *no*, *ni* or *ga* is more frequent in the corpus data than unsuffixed forms, and could therefore be a licit base.

It should be noted, however, that the omission of a case particle is more frequent in informal speech than in writing. Since the NPCMJ is based on written Japanese, it is possible that the relative frequency of dropped particles is higher in spoken Japanese than in the corpus. If unsuffixed nouns were even slightly more frequent in speech than in the corpus, unsuffixed nouns could become relatively more frequent than any one of the suffixed forms. If this were the case, the frequency analysis would, in order to accurately represent the facts, necessarily have to take the cumulative frequency of all suffixed forms into account, rather than that of any one suffixed form. In this event, the vowel duration of the base could be computed as, for example, the mean vowel duration across all suffixed forms.

3.3.3 *Maximal informativeness.* A final method for selecting a base is to choose the base which is 'most informative' (Albright 2002a, b). Under this approach, the base for a given paradigm should be the form which preserves the most contrasts and which allows for accurate generation of as many members of the paradigm as possible. Phonological neutralisations obscure underlying contrasts, therefore forms which undergo neutralisation may be less informative than forms which do not. I argue in the rest of this section that this principle is evident in the case of monomoraic lengthening – monomoraic nouns with case particles are more informative than unsuffixed monomoraic nouns.

3.3.3.1 *Vowel length and the informativeness of incompletely neutralised contrasts*. Albright (2002a: 6) argues that the maximally informative form is the one that 'suffer[s] from the fewest phonological neutralisations, and maintain[s] the most contrasts'. Intuitively, the reduced contrasts of incomplete neutralisation should be less informative than a fully maintained contrast. In the minimal generalisation learner formulated in Albright (2002b), the reliability of a grammatical rule is defined as the number of input forms that the rule accurately derives, divided by the number of forms in which the rule could potentially apply. While this measure considers categorical phonological alternations, the definition of a rule's reliability is not impacted by degree of neutralisation or perceptibility of a contrast. I argue that forms displaying incomplete neutralisation should be considered less informative than those which maintain a complete contrast, but more informative than those in which a contrast is completely neutralised.

In the Japanese case, therefore, incomplete neutralisation of the vowel-length contrast provides reduced reliability and informativeness as compared to forms without any vowel-length neutralisation, and thus might be considered less informative. As such, monomoraic nouns with case particles make better bases than monomoraic nouns without case particles.

If we adopt this analysis, it is crucial to determine whether relative degree of neutralisation factors into determining a form's informativeness. In monomoraic vowel lengthening, the incompletely neutralised contrast yields a surface distinction on the order of 30 ms. This is larger than in some other cases of incomplete neutralisation, such as German final devoicing (approximately 15 ms in the duration of preceding vowels, as noted in §1.1) and flapping in American English (roughly 6 ms; Herd *et al*. 2010, Braver 2014). While we might expect that incompletely neutralised forms in German and English are less informative than those in Japanese, I assume here that neutralised monomoraic nouns are still less informative than non-neutralised forms, and thus that monomoraic nouns with particles (which don't neutralise the length contrast) are better bases than those without particles (which display incomplete neutralisation of the length contrast).

A second neutralisation may further indicate that monomoraic nouns without particles provide less information than those with particles: neutralisation of pitch-accent.

3.3.3.2 *Japanese pitch-accent neutralisation*. Another locus of neutralisation is Japanese pitch accent. In Standard Japanese, phonological words (which may include, for example, a noun and its following particle) may have a pitch accent, usually realised as high tone on the accented mora, followed by a drop in pitch on all following moras in the phrase. (In words of two or more moras, if no accent is present in the first two moras, these moras usually show an LH pattern.) Consider the minimal pairs in (9), from Kawahara (2015).

(9) a. *unaccented* ame+ga LHH 'candy-NOM'
accented a'me+ga HLL 'rain-NOM'
b. *initial accent* ka'ta+ga HLL 'shoulder-NOM'
peninitial accent kata'+ga LHL 'shoulder-NOM'

In phrases ending with short syllables,[9] it is sometimes impossible to determine whether the final mora is accented or unaccented: if a word is recited in isolation (with no carrier sentence), as in (10), from McCawley (1968), the presence or absence of a final accent is, for most speakers, not apparent.[10]

(10) *final accent* atama' LHH 'head'
unaccented miyako LHH 'city'

This accentual ambiguity is remedied if the word is followed by another mora, such as is provided by a case particle – the following mora surfaces as low if the word has a final accent, but as high if the word does not, as in (11).

(11) atama'+ga LHHL 'head-NOM'
miyako+ga LHHH 'city-NOM'

By definition, monomoraic words will end (and, indeed, begin) with a short syllable, making them subject to accent neutralisation in phrase-final position. The neutralisation of tone in isolated monomoraic words is shown in (12a), with the contrast being realised when a case particle is present, as in (12b).[11]

[9] In phrases ending with long syllables, the entire HL of the pitch accent can be realised.

[10] Several previous experiments indicate that while some speakers completely neutralise this contrast, others may neutralise only incompletely (Sugito 1982, Poser 1984, Pierrehumbert & Beckman 1988, Kubozono 1993, Vance 1995). Vance (1995), for example, found in one experiment that eleven out of 14 participants showed no clear F0 difference between final accented and unaccented moras. These studies also provide mixed evidence that speakers are more likely to completely neutralise when the target word is pronounced in isolation, without following phonological material (e.g. not in a carrier sentence). This might explain why Fig. 1c, which contains a noun in a carrier sentence, seems to show an HL pattern. Even in cases where a small distinction is maintained between accented and unaccented final syllables, the incompletely neutralised contrast is still weaker evidence than a full contrast, and thus might be considered less informative than the full-fledged contrasts in non-word-final position, as argued above for vowel duration.

[11] Words consisting of a single mora, when produced in isolation, surface with just an H tone, contrary to the generalisation for longer words which, if unaccented, surface with an initial LH (McCawley 1968: 133, n.14, Vance 1987).

(12) a.	*accented*	ki' H	'tree'
	unaccented	ki H	'spirit'
b.	*accented*	ki'+ga HL	'tree-NOM'
	unaccented	ki+ga HH	'spirit-NOM'

While it is true that a mora may be supplied by any following word in the phonological phrase, nouns in isolation lacking case particles are clearly less informative than those with case particles as regards final pitch accent. This suggests that nouns with case particles may be better bases than those without – the desired result – on the basis of informativeness.

3.3.4 *Summary of basehood.* Given these four approaches to choosing a base, morphological complexity and orthography are least applicable to the case of Japanese monomoraic noun lengthening. While both frequency and informativeness can produce the desired result – namely, that the base be a monomoraic noun with a particle – I will assume the informativeness definition throughout.

One reason for this assumption stems from the concern about the cumulative *vs*. individual frequency of particles. If a frequency-based approach is chosen, the corpus data suggests that any of *o*, *no*, *ni* or *wa* could be the particle of choice for the base. As was noted above, however, the corpus data comes from written, rather than spoken, Japanese. If the proportion of particle-less nouns (11.66%) were to rise even slightly in spoken contexts, such forms could overtake any individual particle (e.g. *o*, the most frequent particle in the corpus, was found with 17.11% of nouns). This would force a definition of frequency-based basehood that groups all forms with particles together, creating a sort of meta-base – a complication that the informativeness-based approach avoids.

4 The Weighted Paradigm Uniformity theory

In this section I introduce the Weighted Paradigm Uniformity (WPU) theory of incomplete neutralisation. As described in §3.2, the WPU model combines two independently motivated theoretical mechanisms: paradigm uniformity and weighted phonetic constraints.

4.1 Duration targets for monomoraic and bimoraic nouns

I assume here that, for any given segment, there is some language-specific target duration. For the case of Japanese, I propose two cover targets that govern the duration of all monomoraic and bimoraic segments: all segments which bear one mora in the output have a target duration represented as *TargetDur*(μ), and all segments which bear two moras in the output have a target duration represented as *TargetDur*($\mu\mu$). Using the rounded values from Braver & Kawahara (2016), shown in Table I, unlengthened monomoraic short vowels have *TargetDur*(μ) = 50 ms and underlyingly bimoraic long vowels have $TargetDur(\mu\mu) = 150$ ms. In non-neutralising contexts, both of these targets are generally met.

Along the lines of Flemming's (2001) C-DURATION and σ-DURATION constraints, I assume that candidates are under pressure to conform to language-specific duration targets. This pressure is codified in two constraints of the family DUR(x)=TARGETDUR(x). DUR(μ)=TARGETDUR(μ), defined in (13), penalises monomoraic candidates which fail to match their target duration. In a parallel fashion, DUR($\mu\mu$)=TARGETDUR($\mu\mu$) penalises bimoraic candidates which do not meet their target duration.[12]

(13) DUR(μ)=TARGETDUR(μ)

The duration of a mora-bearing unit which bears a single mora in the output should match the target (canonical) output duration for that mora-bearing unit when it bears one mora.

Formal definition

For a mora-bearing unit α which bears one mora in the output, and is spoken at speech rate $\mathbb{R}$, let:

- $TargetDur(\mu)$ be the canonical output duration of α when bearing one mora in the output, and spoken at speech rate $\mathbb{R}$
- $Dur(cand)$ be the actual duration of α under evaluation, spoken at speech rate $\mathbb{R}$
- w_μ be the weight applied to this constraint
- the total cost assessed by DUR(μ)=TARGETDUR(μ) be $cost = w_\mu(TargetDur(\mu) - Dur(cand))^2$

To see DUR(μ)=TARGETDUR(μ) in action, consider the tableau in (14) (which assumes a constraint weight of 1).

(14)

/ki mo/ (unlengthened short)	DUR(μ)=TARGETDUR(μ)
a. V duration = 30 ms	400 $(1 \times (50 - 30)^2)$
b. V duration = 40 ms	100 $(1 \times (50 - 40)^2)$
☞ c. V duration = 50 ms	0 $(1 \times (50 - 50)^2)$
d. V duration = 60 ms	100 $(1 \times (50 - 60)^2)$
e. V duration = 70 ms	400 $(1 \times (50 - 70)^2)$

The DUR(μ)=TARGETDUR(μ) column shows the cost of each candidate for that constraint; the parentheses show the calculation used to reach that cost. For example, in candidate (14a), the cost associated with DUR(μ)=TARGETDUR(μ) is calculated as in (15), where the candidate's

12 The cost function for DUR(μ)=TARGETDUR(μ) and the following constraints square the difference between the target and the actual duration for two reasons. First, I assume that a candidate cannot have a negative cost; squaring the difference ensures that all costs assessed are positive. Second, doing so creates a quadratic equation, which makes it possible to mathematically compute the minimum value for a cost function.

duration (*Dur*(*cand*)) is 30 ms and the target duration for a single mora (as per the rounded data) is 50 ms.

(15) $cost = w_\mu(TargetDur(\mu) - Dur(cand))^2$
$= 1 \times (50 - 30)^2$
$= 400$

As can be seen in (14), DUR(μ)=TARGETDUR(μ) penalises short vowels whose duration differs from the target (50 ms); as candidates' durations diverge from this target, the cost increases.

DUR(μμ)=TARGETDUR(μμ) works nearly identically, with two moras instead of one. As can be seen in the tableau in (16), candidates with surface-bimoraic vowels (both underlyingly long vowels as in (16a) and lengthened short vowels as in (b)) decrease in cost for DUR(μμ)=TARGETDUR(μμ) as they approach *TargetDur*(μμ) (150 ms).

(16)

		DUR(μμ)=TARGDUR(μμ)	FTBIN(μ)	DEP(μ)	*total*
a.	/kii (μμ)/				
	i. [μμ] V dur=130 ms	400 $(1 \times (150 - 130)^2)$	0	0	400
	ii. [μμ] V dur=140 ms	100 $(1 \times (150 - 140)^2)$	0	0	100
	☞ iii. [μμ] V dur=150 ms	0 $(1 \times (150 - 150)^2)$	0	0	0
	iv. [μ] V dur=150 ms	0 (*vacuous*)	2	0	2
	v. [μμ] V dur=160 ms	100 $(1 \times (150 - 160)^2)$	0	0	100
	vi. [μμ] V dur=170 ms	400 $(1 \times (150 - 170)^2)$	0	0	400
b.	/ki ∅ (μ)/				
	i. [μμ] V dur=130 ms	400 $(1 \times (150 - 130)^2)$	0	1	401
	ii. [μμ] V dur=140 ms	100 $(1 \times (150 - 140)^2)$	0	1	101
	☞ iii. [μμ] V dur=150 ms	0 $(1 \times (150 - 150)^2)$	0	1	1
	iv. [μ] V dur=150 ms	0 (*vacuous*)	2	0	2
	v. [μμ] V dur=160 ms	100 $(1 \times (150 - 160)^2)$	0	1	101
	vi. [μμ] V dur=170 ms	400 $(1 \times (150 - 170)^2)$	0	1	401

In order to prevent candidates which fail to epenthesise a second mora, and thus vacuously satisfy DUR(μμ)=TARGETDUR(μμ), the tableau in (16) also contains two additional constraints not yet discussed: FTBIN(μ) prefers feet to be minimally bimoraic (with a cost of 1 per monomoraic foot), and DEP(μ) assesses a cost of 1 per epenthesised mora. When FTBIN(μ) is weighted above DEP(μ) (here $w_{\mathrm{FtBin}(\mu)} = 2$, $w_{\mathrm{Dep}(\mu)} = 1$), candidates like (16b.iv) cannot win by failing to epenthesise and thus vacuously fulfilling DUR(μμ)=TARGETDUR(μμ). In all future tableaux and cost tabulations, I exclude monomoraic candidates which fail to lengthen, on the assumption that a weighting in which FTBIN(μ) outweighs DEP(μ) prevents such candidates from being viable optima.

4.2 Paradigm uniformity in lengthened monomoraic nouns

Given the constraints DUR(μ)=TARGETDUR(μ) and DUR($\mu\mu$)=TARGETDUR ($\mu\mu$), it is clear how unlengthened short vowels and underlyingly long vowels are under pressure to reach their duration targets. We are left, however, with vowels in lengthened short nouns, which, with an average duration of 125 ms, meet neither the target for short nouns (50 ms) nor the target for long nouns (150 ms). Because lengthened short nouns are surface-bimoraic, they are subject to the constraint DUR($\mu\mu$)=TARGETDUR($\mu\mu$); note, however, that they do not comply completely with this constraint. I claim that the distinction between underlyingly long nouns on the one hand and lengthened short nouns on the other is caused by pressure on lengthened short nouns to maintain a degree of faithfulness to canonical short vowels – a pressure that underlyingly long vowels do not face. To see how this works, we will first examine a single example based on the rounded data in Table I, and then proceed to a more formal definition.

Consider the three utterances in (17), which contain (a) unlengthened short vowels, (b) lengthened short vowels and (c) underlyingly long vowels. Vowel durations of the underlined words are given in parentheses.[13]

(17)	a.	[ki mo]$_\omega$	nakushita	yo		(50 ms)
		tree PRT	found	DISC		
	b.	[ki Ø]$_\omega$	nakushita	yo		(125 ms)
		tree	found	DISC		
	c.	[kii Ø]$_\omega$	nakushita	yo		(150 ms)
		key	found	DISC		

The vowel in (17b) has two surface moras (due to lengthening), and is thus pressured by DUR($\mu\mu$)=TARGETDUR($\mu\mu$) to reach a target of 150 ms. It does not, however, reach this target. This failure is due to pressure from a competing constraint, which requires faithfulness to a morphologically related base form (i.e. paradigm uniformity).

As per the discussion in § 3.3, we will take the base to which monomoraic nouns must be faithful to be a form which contains the noun and a following case particle. I formalise this faithfulness to a morphologically related base in the form of an output–output constraint, OO-ID(dur), which enforces faithfulness between a given segment in a candidate and the corresponding segment in the candidate's base, as in (18).

[13] PRT = comitative particle, DISC = discourse marker.

(18) OO-ID(dur)

The duration of a segment in the candidate should be faithful to the duration of the same segment in the base.

Formal definition

For a segment α in the candidate, let:

- *Dur*(*cand*) be the duration of α in the output
- *Dur*(*base*) be the duration of the segment β in the base that corresponds to α in the candidate (i.e. $(\alpha, \beta) \in \mathcal{R}_{OO}$)
- w_{ID} be the weight applied to this constraint
- the total cost assessed by OO-ID(dur) be
  $cost = w_{\mathrm{ID}}(Dur(cand) - Dur(base))^2$

To see this more concretely, let us apply OO-ID(dur) to the lengthened [ki Ø]$_\omega$ in (17b). As stated above, the base to which (17b) must be faithful has a vowel duration of 50 ms. OO-ID(dur) therefore pressures [ki Ø]$_\omega$ to attain a vowel duration of 50 ms.

The effect of OO-ID(dur) on lengthened short nouns can be seen in the tableau in (19): the cost of this constraint increases as vowel durations move away from 50 ms.

(19)

/ki Ø/$_\omega$ (lengthened short)	OO-ID(dur)
a. V duration = 25 ms	625 $(1 \times (25-50)^2)$
☞ b. V duration = 50 ms	0 $(1 \times (50-50)^2)$
c. V duration = 75 ms	625 $(1 \times (75-50)^2)$
d. V duration = 100 ms	2500 $(1 \times (100-50)^2)$

4.3 Interaction of OO-ID(dur) and DUR($\mu\mu$)=TARGETDUR($\mu\mu$)

We consider now the tension between paradigm uniformity and the canonical duration of a segment – in other words, between OO-ID(dur) and DUR ($\mu\mu$)=TARGETDUR($\mu\mu$). It should be noted that DUR(μ)=TARGETDUR(μ) is not relevant for assessing lengthened short vowel inputs, because it only penalises mora-bearing units with a single mora on the surface. Since lengthened short candidates by definition have two moras on the surface, DUR(μ)=TARGETDUR(μ) does not apply, and will not be shown in the tableaux that follow. Similarly, as discussed above, the relative weighting of FTBIN(μ) over DEP(μ) prevents lengthened nouns from surfacing with only one mora; the costs of these two constraints will be left out of the calculations below, to simplify the exposition.

The key tension between DUR($\mu\mu$)=TARGETDUR($\mu\mu$) and OO-ID(dur) can be expressed in terms of their role in setting target durations for lengthened short candidates. DUR($\mu\mu$)=TARGETDUR($\mu\mu$) is least costly for lengthened short candidates whose durations approach *TargetDur*($\mu\mu$) (150 ms), while OO-ID(dur) is least costly for those candidates whose durations approach *TargetDur*(μ) (50 ms), since their base has only one mora.

The relative strength of these constraints – as reflected in their weighting – determines how the conflict between these two pressures is resolved. As the weight of DUR($\mu\mu$)=TARGETDUR($\mu\mu$) increases relative to that of OO-ID(dur), the candidates with the lowest cost are those that approach 150 ms. Conversely, as the weight of OO-ID(dur) increases relative to that of DUR($\mu\mu$)=TARGETDUR($\mu\mu$), the candidates with the lowest cost are those that approach 50 ms.

The degree to which a candidate satisfies the compromise between DUR ($\mu\mu$)=TARGETDUR($\mu\mu$) and OO-ID(dur) can be assessed by its total cost – the sum of the costs of all constraints for that candidate. The candidate which best satisfies the compromise (and is therefore the winner) is the one which has the lowest overall cost. Given our two constraints, the total cost of a candidate is as in (20).

(20) *Total cost* = *cost*(OO-ID(dur)) + *cost*(DUR($\mu\mu$)=TARGETDUR($\mu\mu$))

As shown in Table IV, with the weighting <$w_{\mathrm{Id}} = 1$, $w_{\mu\mu} = 3$> (and with the same assumptions about durational targets as above), the total cost of a candidate decreases as the duration of lengthened vowels approaches 125 ms – the minimum cost is attained by candidates of 125 ms (our desired winner).

Lengthened V dur (ms)	*cost*(OO-ID(dur)) $w_{\mathrm{ID}}(Dur(cand) - Dur(base))^2$	*cost*(DUR($\mu\mu$)=TARGDUR($\mu\mu$)) $w_{\mu\mu}(TargDur(\mu\mu) - Dur(\mu\mu))^2$	*total cost*
75	$1 \times (75 - 50)^2$	$3 \times (150 - 75)^2$	17500
100	$1 \times (100 - 50)^2$	$3 \times (150 - 100)^2$	10000
125	$1 \times (125 - 50)^2$	$3 \times (150 - 125)^2$	7500
150	$1 \times (150 - 50)^2$	$3 \times (150 - 150)^2$	10000

Table IV

Costs for given lengthened short vowel durations, where $w_{\mathrm{Id}} = 1$, $w_{\mu\mu} = 3$, $TargetDur(\mu) = 50$ ms and $TargetDur(\mu\mu) = 150$ ms.

Varying the weights of these two constraints results in a different minimum to the cost function, and a different degree of neutralisation, as might be expected to be found in other languages, as shown in Table V. As the weight of w_{Id} increases relative to that of $w_{\mu\mu}$, the duration of the lengthened vowels is predicted to decrease (and, *vice versa*, to increase). This principle can also be seen in the graph in Fig. 2, in which the x- and y-axes represent the weights of OO-ID(dur) and DUR ($\mu\mu$)=TARGETDUR($\mu\mu$) ($w_{\mathrm{Id}} = x$; $w_{\mu\mu} = y$; $0 < x$, $y \geq 5$), and the z-axis represents the predicted duration of lengthened vowels: the WPU model can generate languages anywhere along the plotted surface.

w_{Id}	$w_{\mu\mu}$	lengthened vowel duration
2	1	83.33
1	1	100.00
1	2	116.17
1	3	125.00
1	4	130.00

Table V

Predicted lengthened short vowel durations (in ms) for sample weightings of OO-ID(Dur) (w_{Id}) and DUR($\mu\mu$)=TARGETDUR($\mu\mu$) ($w_{\mu\mu}$), where $TargetDur(\mu) = 50$ ms and $TargetDur(\mu\mu) = 150$ ms.

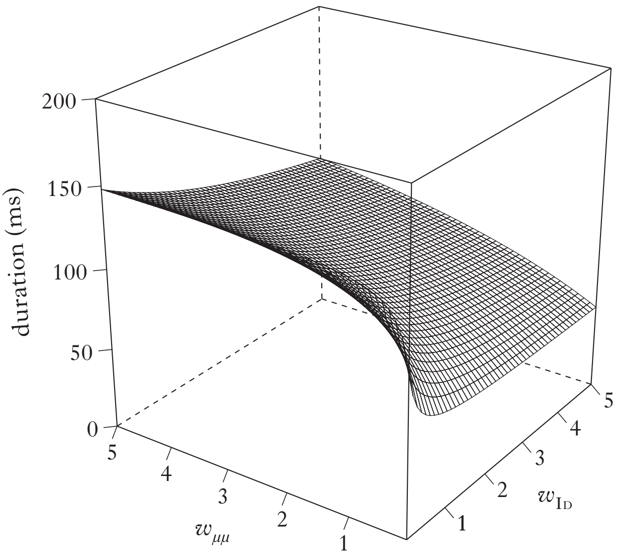


Figure 2

Predicted lengthened vowel duration, given weights for w_{Id} and $w_{\mu\mu}$ (TARGETDUR(μ)=50 ms; TARGETDUR($\mu\mu$)=150 ms).

5 Discussion

The preceding sections have laid out the case of Japanese monomoraic lengthening and its analysis in the WPU model. In this section I outline the relation between the WPU model and other theories of incomplete neutralisation, and discuss their typological predictions.

5.1 Relation to prototype and exemplar models

It is worthwhile to note that the notions of canonical realisation and bases determined by frequency are, in many ways, reminiscent of prototype models (Posner & Keele 1968, Reed 1972, Rosch 1973, Smith & Minda 2000) and exemplar models (Nosofsky 1986, Lacerda 1995, Goldinger 1996, Johnson 1997, Pierrehumbert 2001a, *inter alia*). Such models and their close kin have been applied to the study of incomplete neutralisation. For example, Yu (2007) gives a clear account of a case of near-merger in Cantonese tone in terms of an exemplar-based model. On this view, near-merger occurs when two or more exemplar clouds begin to overlap, but category membership remains distinct.

Relatedly, Kirby (2010) uses a computational model with a number of similarities to exemplar models to analyse incomplete neutralisation in Dutch word-final obstruents (as described by Warner *et al.* 2004). This work shows that when only a single acoustic cue to incomplete neutralisation is considered, complete neutralisation may be predicted, but that as more cues are added to the model, incomplete neutralisation is more likely. While the WPU model developed here is not formally related to prototype or exemplar models, it does share the idea of 'typicality'. The WPU model, then, serves to formalise these properties of prototype and exemplar models in the language of a more traditional generative framework. As pointed out by an anonymous reviewer, there are few, if any, exemplar models of incomplete neutralisation that are sufficiently worked out to produce quantitative predictions, though nothing about such models is inherently incompatible with them. The WPU model is offered, then, as an example of the manner in which quantitative predictions can be made in a framework that exploits the notion of typicality.

5.2 Typological predictions of the WPU model of incomplete neutralisation

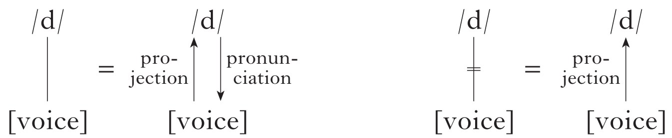
The WPU model of incomplete neutralisation described in this paper makes four predictions: (i) incomplete neutralisation can occur even when no underlying association between a segment and a feature (or prosodic unit) has been delinked, (ii) there can be gradient degrees of neutralisation, (iii) morphological and/or phonetic information can influence incomplete neutralisation and (iv) incomplete neutralisation can occur even when only one relevant process is at play.

5.2.1 *Lack of underlying associations.* The WPU model of incomplete neutralisation predicts that incomplete neutralisation can occur in processes that do not result, in autosegmental terms, in the delinking of a feature (or prosodic unit) from a segment. This contrasts with Turbidity Theory, which allows incomplete neutralisation only when such a delinking occurs.

As an example, the turbidity model of incomplete neutralisation, as applied to final devoicing, argues that underlyingly voiceless and devoiced segments have distinct representations at the level of the phonological output. More specifically, underlyingly voiceless segments have neither a projection nor a pronunciation relation with a [voice] feature (and thus surface as voiceless), whereas devoiced segments have a projection relation only. In van Oostendorp's (2008) formulation, incomplete neutralisation results when this lone projection relation plays a role in phonetic interpretation – devoiced segments will be coloured by the presence of the projection relation with [voice], even though they lack a pronunciation relation with [voice].

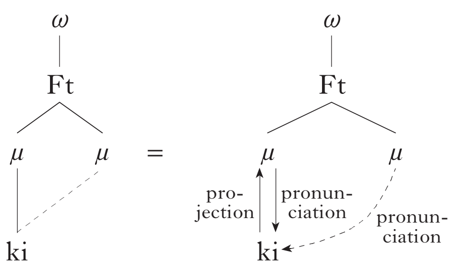
The projection relation, Goldrick (2000: 2–3) suggests, is akin to a delinked association line: in final devoicing, devoiced segments originate with an association line to a [voice] feature, but this line is cancelled, leaving only an abstract, structural relation, as in (21b). Goldrick also suggests that pronunciation relations are similarly equivalent to (dotted) association lines (as will be seen in (22)).

(21) a. *Voiced segments* b. *Devoiced segments*



In essence, projection marks the segment under discussion as having underlyingly been associated with a [voice] feature. The turbidity approach to incomplete neutralisation capitalises on the fact that, in this model, underlying voicing status is available to the phonetic module (via the presence/absence of a projection relation). Incomplete neutralisation, then, occurs where an underlying association between a segment and a feature (or prosodic unit) has been delinked – or, in turbidity terms, originates with both types of relations but ends with only a projection relation. The case of Japanese monomoraic lengthening, where no such delinking occurs, does not follow this principle: rather than a lone projection relation causing incomplete neutralisation, in Japanese the link between a second mora and a lengthened segment consists of a pronunciation relation, as in (22).

(22)



One option for turbidity theorists wishing to allow this sort of incomplete neutralisation would be to add a stipulation that the phonetic implementation of moras linked only by a pronunciation relation should be different from that of moras linked by both relations. This potential remedy does not, however, come without a cost. Normally, segments with a production relation to a feature should exhibit a canonical manifestation of that feature (e.g. in the case of (de)voicing, /d/ → [d] has a pronunciation relation with [voice], and is canonically pronounced as voiced). In the Japanese case, the addition of a pronunciation relation to the second mora does not result in the canonical manifestation of a second mora (i.e. it is not as long). Allowing non-canonical manifestations of features related by pronunciation would seem to change the predictions of Turbidity Theory in general, allowing a new kind of 'unfaithful' pronunciation.

The case of incomplete neutralisation in American English flapping also presents a problem for the turbidity approach, due to its requirement for underlying associations: the realisation of segments with only a projection relation to [voice] would differ significantly between flapping and final devoicing. In final devoicing, as in (21), devoiced segments are related to [voice] by a projection relation. Such segments are pronounced more like voiceless segments than voiced ones – they have shorter preceding vowels, less voicing into closure and more aspiration (Port & O'Dell 1985). Flaps deriving from /t/, however, have no underlying relation to [voice]; they must gain a pronunciation relation to [voice] during the derivation. These segments, unlike the devoiced segments, are pronounced more like voiced segments than voiceless ones. This suggests that segments with only a projection relation to [voice] must be realised differently in different languages and/or circumstances.

Consequently, the WPU model predicts incomplete neutralisation both in cases where underlying associations exist (e.g. devoicing) and in cases where they don't (e.g. Japanese monomoraic lengthening), while the turbidity approach would require additional stipulations to account for the latter type.

5.2.2 *Gradient degrees of neutralisation*. The WPU approach to incomplete neutralisation, in contrast to the OT-CC model (Gouskova & Hall 2009), predicts a relatively unrestricted set of possible 'degrees' of neutralisation (for example, where a segment lies on a continuum between voiced and voiceless, or how long a lengthened vowel is relative to underlyingly short *vs*. long vowels). OT-CC's limits on possible degrees of neutralisation follow from two assumptions of the theory: (i) constraints refer to phonological, rather than phonetic units, and (ii) steps along the chain represent, at least in most formulations, a change that incurs exactly one violation of a faithfulness constraint. This contrasts with the WPU model, which allows for a gradient degree of neutralisation.

In the OT-CC model, incomplete neutralisation results when the phonetic module chooses an intermediate stage of the candidate chain, rather than the final step of the chain. In Lebanese Arabic, then, epenthesis of [i] actually results from a chain of changes: <CC, CɨC, CəC, CiC>, where each step in the chain is a step up the sonority hierarchy. Speakers who incompletely neutralise this contrast produce a form from the middle of the chain.

It is not clear, however, what intermediate step would be relevant in monomoraic lengthening. One approach would be to assume that the insertion of a mora and its linking to a segment represent two steps along the chain – in this way, a stray unlinked mora might be interpreted by the phonetic module in such a manner that it results in a slight degree of lengthening. This approach, though, seems to come up against a problem: assuming that the Japanese bimoraicity requirement is only satisfied by moras that are linked to segmental content, the insertion of a second (empty) mora in a short noun is not a harmonically improving step.

A similar issue arises with incomplete neutralisation of final devoicing: there are, under most views, only two values for voicing: voiced or voiceless. As such, no intermediate steps between voicing and voicelessness are created along the chain – and there is therefore no intermediate step to select for incompletely neutralised forms. A plausible resolution might be to assume that there is indeed a third value for voicing features – perhaps a candidate underspecified for voice intervenes between voiced and voiceless segments. I leave aside here the debate regarding non-binary features, but note that it is not obvious that segments underspecified for voice should be produced as incompletely neutralised.

A potential remedy to this problem would be to allow OT-CC constraints to refer directly to phonetic values, with a concomitant stipulation that steps along a candidate chain can differ only to some particular degree along a phonetic continuum – perhaps being restricted to steps of a just noticeable difference. Indeed, McCarthy (2009) discusses reference to phonetic details, as encoded in the P-map (Steriade 2009), in Harmonic Serialism; however no such claims seem to have been explored with respect to OT-CC, and the typological properties of such a system are not known. Setting aside the precise details of such a proposal, one might imagine a candidate chain for vowel lengthening to consist of vowels of increasing duration: <Ø, V[10 ms], V[20 ms], V[30 ms], …>. If a mechanism for choosing the correct step along the chain could be developed (a non-trivial task), vowels of intermediate durations might be selected in cases of incomplete neutralisation.

Applying this remedy to cases of final devoicing appears to be significantly more complicated than for vowel duration. While vowel duration is easily quantified in a single measurement (i.e. duration itself), voicing contrasts are manifested across a wide variety of acoustic cues, including duration of the preceding vowel (Chen 1970), closure duration (Kluender *et al*. 1988) and F0 and F1 (Hombert *et al*. 1979, Kingston & Diehl 1994). Organising these properties into clear steps along a single continuum – while ensuring that each step along the chain increases in harmony – seems like a task that is not sufficiently motivated by the incomplete neutralisation data alone. The WPU model of incomplete neutralisation, in contrast to OT-CC, predicts cases of incomplete neutralisation where the neutralised segment is not clearly defined by a single violation of a faithfulness constraint, as in the Japanese case and in final devoicing.

5.2.3 *Reference to phonetic and paradigmatic information.* The WPU model of incomplete neutralisation, by virtue of its dependence on phonetic and paradigmatic information, assumes that such information can influence incomplete neutralisation. In contrast, the turbidity approach predicts that only phonological information should play a role in determining incomplete neutralisation. This is because turbidity relies on relations with an underlying feature or prosodic unit, neither of which is generally assumed to provide information about the phonetic implementation of the relevant segment or about the realisation of paradigm members. Two sorts of incomplete neutralisation processes could help to determine whether this prediction is desirable: (i) processes that necessarily refer to phonetic detail, rather than phonological structure, and (ii) processes that clearly map to a morphological relative rather than an underlying form.

5.2.4 *Incomplete neutralisation in single processes.* The WPU model predicts that incomplete neutralisation can occur when only one relevant process occurs. In the Japanese case, the only change involves the duration of a vowel. This contrasts with cases of final devoicing and flapping in which these processes are accompanied by the additional process of vowel lengthening.

Under Anderson's (1975) rule-ordering approach to incomplete neutralisation, a flapping rule in American English interacts with a rule that lengthens vowels before voiced segments. Crucially, the vowel-lengthening rule must apply first, since the flapping rule feeds lengthening by changing voiceless /t/ to a (voiced) flap. Because the vowel-lengthening rule is gradient, while the flapping rule is categorical, Anderson argues that some gradient phonetic rules (like vowel lengthening) must be allowed to precede some categorical rules (like flapping). The prediction of this theory is that any pair of categorical and gradient rules can be interleaved, resulting in otherwise opaque rule application. This analysis can be extended to cases of final devoicing – a gradient vowel lengthening rule (e.g. $V \rightarrow V^{\cdot} / \_\_ \; C_{[+\text{voice}]}$) could apply before a categorical rule of obstruent devoicing, yielding longer vowels before devoiced segments than before underlyingly voiceless segments.

It is unclear, however, how such an analysis would apply to the Japanese monomoraic lengthening process described above, since there is only one rule at play ($/\mu/ \rightarrow \mu\mu \; / \; [\,\_\_\,]_{\omega}$). If this lengthening rule is gradient, we should not expect the lengthened vowels to fulfil the categorical bimoraicity requirement. If we instead assume that this rule is categorical, we do not predict that lengthened vowels should be intermediate in length. Flapping and final devoicing avoid this issue, because two distinct processes are involved – e.g. vowel lengthening and flapping, or vowel lengthening and final devoicing. As such, rules for each of these processes can be interleaved, allowing a gradient process to precede a phonological one. In the Japanese case, however, only one process is involved: vowel lengthening is either categorical or gradient – it can't be both. More generally, the rule-ordering approach predicts incomplete neutralisation only in cases where multiple processes are interacting.

6 Conclusion

6.1 The functional explanation for incomplete neutralisation

A functional explanation for incomplete neutralisation is proposed by Port & Crawford (1989), who argue from experimental evidence that speakers vary the completeness of incompletely neutralised contrasts depending on the communicative situation. This result mirrors those from other phenomena in which speakers manipulate their productions for the benefit of hearers (see e.g. Lindblom 1990, Scarborough 2003, 2010, Flemming 2010, Syrett & Kawahara 2013). On this view, speakers maintain a small trace of an underlying contrast, even in neutralising positions, to aid the listener in comprehending what is said.

This functional explanation provides a clear motivation for cases of incomplete neutralisation with a relatively large surface distinction: maintaining a contrast is worthwhile, as long as the listener can perceive it. Less clear, though, is how this idea explains cases of incomplete neutralisation with magnitudes (on average) so small as to be imperceptible. This includes both individual speakers with small contrasts and processes which maintain a small contrast across multiple speakers. Some processes which result in incomplete neutralisation have shown consistently small contrasts: for example, Braver (2014) found that the distinction between vowels preceding /d/-flaps and those preceding /t/-flaps was on average only 5.69 ms (the speaker with the largest distinction showed a 15.28 ms difference). Estimates for the just noticeable difference (JND) in vowel length vary, but generally range between 10 and 25 ms (Fujisaki *et al.* 1975, Klatt & Cooper 1975, Nooteboom & Doodeman 1980), suggesting that vowel duration in incomplete neutralisation of English flapping may be below the threshold for discrimination, while the Japanese data from Braver & Kawahara (2016) might just exceed the highest end of that JND range.

In terms of the WPU model, speakers would be weighting OO-Id(dur) high enough to cause a surface trace of the underlying contrast, but at the same time so low that listeners cannot perceive that trace. In spite of the growing body of evidence that speakers have deep knowledge of the perceptibility of contrasts (e.g. Steriade 2009), it would seem that this knowledge may not always be involved in calculating the weighting of OO-Id(dur) in incomplete neutralisation.

6.2 Final remarks

The WPU model of incomplete neutralisation makes use of two independently motivated grammatical mechanisms: weighted phonetic constraints and paradigm uniformity. By casting incomplete neutralisation in terms of weighted phonetic constraints, the WPU model pits faithfulness to a morphologically related base against the canonical target duration of non-neutralised segments, generating quantitative predictions about incompletely neutralised contrasts. The model also makes typological predictions that are distinct from existing theories of incomplete neutralisation, as shown by the ability of the WPU model to capture the Japanese monomoraic lengthening data.

REFERENCES

Albright, Adam (2002a). Base selection in analogical change in Yiddish. *BLS* **28**. 1–13.

Albright, Adam (2002b). *The identification of bases in morphological paradigms*. PhD dissertation, University of California, Los Angeles.

Anderson, Stephen R. (1975). On the interaction of phonological rules of various types. *JL* **11**. 39–62.

Barnes, Jonathan (2006). *Strength and weakness at the interface: positional neutralization in phonetics and phonology*. Berlin & New York: Mouton de Gruyter.

Behne, Dawn, Takayuki Arai, Peter Czigler & Kirk Sullivan (1999). Vowel duration and spectra as perceptual cues to vowel quantity: a comparison of Japanese and Swedish. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley: University of California. 857–860.

Benua, Laura (1997). *Transderivational identity: phonological relations between words*. PhD dissertation, University of Massachusetts, Amherst.

Bermúdez-Otero, Ricardo (2007). Diachronic phonology. In Paul de Lacy (ed.) *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press. 497–517.

Braver, Aaron (2013). *Degrees of incompleteness in neutralization: paradigm uniformity in a phonetics with weighted constraints*. PhD dissertation, Rutgers University.

Braver, Aaron (2014). Imperceptible incomplete neutralization: production, non-identifiability, and non-discriminability in American English flapping. *Lingua* **152**. 24–44.

Braver, Aaron & Shigeto Kawahara (2014). Incomplete vowel lengthening in Japanese: a first study. *WCCFL* **31**. 86–95.

Braver, Aaron & Shigeto Kawahara (2016). Incomplete neutralization in Japanese monomoraic lengthening. In Adam Albright & Michelle A. Fullwood (eds.) *Proceedings of the 2014 Meeting on Phonology*. http://dx.doi.org/10.3765/amp.v2i0.3757.

Chang, Steve S., John J. Ohala, Gunnar Hansson, Benjamin James, Julie Lewis, Lily Liaw, Margaret Urban, Alan Yu & Kenneth Van Bik (1999). Vowel-dependent VOT variation: an experimental study. *JASA* **105**. 1400.

Chen, Matthew (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* **22**. 129–159.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Dinnsen, Daniel A. & Jan Charles-Luce (1984). Phonological neutralization, phonetic implementation and individual differences. *JPh* **12**. 49–60.

Dinnsen, Daniel A. & M. Garcia-Zamor (1971). The three degrees of vowel length in German. *Papers in Linguistics* **4**. 111–126.

Dmitrieva, Olga (2005). *Incomplete neutralization in Russian final devoicing: acoustic evidence from native speakers and second language learners*. MA thesis, University of Kansas.

Fisher, William M. & Ira J. Hirsh (1976). Intervocalic flapping in English. *CLS* **12**. 183–198.

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

Flemming, Edward (2002). *Auditory representations in phonology*. London & New York: Routledge.

Flemming, Edward (2010). Modeling listeners: comments on Pluymaekers et al. and Scarborough. In Fougeron *et al*. (2010). 587–605.

Fougeron, Cécile, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.) (2010). *Laboratory phonology 10*. Berlin & New York: De Gruyter Mouton.

Fougeron, Cécile & Donca Steriade (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? In *Proceedings of the 5th European Conference on Speech Communication and Technology* (*Eurospeech 1997*). Rhodes, Greece. 943–946. http://www.isca-speech.org/archive/eurospeech_1997.

Fourakis, Marios & Gregory K. Iverson (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica* **41**. 140–149.

Fox, Robert A. & Dale Terbeek (1977). Dental flaps, vowel duration and rule ordering in American English. *JPh* **5**. 27–34.

Fujisaki, Hiroya, Kimie Nakamura & Toshiaki Imoto (1975). Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (eds.) *Auditory analysis and perception of speech*. London: Academic Press. 197–219.

Gerfen, Chip (2002). Andalusian codas. *Probus* **14**. 247–277.

Gianinnia, Antonella & Umberto Cinque (1978). *Phonetic status and phonemic function of the final devoiced stops in Polish*. Naples: Laboratorio di Fonetica Sperimentale.

Goldinger, Stephan D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**. 1166–1183.

Goldrick, Matthew (2000). Turbid output representations and the unity of opacity. *NELS* **30**. 231–245.

Goldrick, Matthew & Sheila E. Blumstein (2006). Cascading activation from phonological planning to articulatory processes: evidence from tongue twisters. *Language and Cognitive Processes* **21**. 649–683.

Gouskova, Maria & Nancy Hall (2009). Acoustics of epenthetic vowels in Lebanese Arabic. In Steve Parker (ed.) *Phonological argumentation: essays on evidence and motivation*. London: Equinox. 203–225.

Haraguchi, Shosuke (1977). *The tone pattern of Japanese: an autosegmental theory of tonology*. Tokyo: Kaitakusha.

Herd, Wendy, Allard Jongman & Joan Sereno (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *JPh* **38**. 504–516.

Hirata, Yukari (2004). Effects of speaking rate on the vowel length distinction in Japanese. *JPh* **32**. 565–589.

Hirata, Yukari & Kimiko Tsukada (2009). Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica* **66**. 129–149.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan (1979). Phonetic explanations for the development of tones. *Lg* **55**. 37–58.

Itô, Junko (1990). Prosodic minimality in Japanese. *CLS* **26:2**. 213–239.

Ito, Junko & Armin Mester (2003). Weak layering and word binarity. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin'ichi Tanaka (eds.) *A new century of phonology and phonological theory: a Festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*. Tokyo: Kaitakusha. 26–65.

Jassem, Wiktor & Lutosława Richter (1989). Neutralization of voicing in Polish obstruents. *JPh* **17**. 317–325.

Johnson, Keith (1997). Speech perception without speaker normalization: an exemplar model. In Keith Johnson & John W. Mullenix (eds.) *Talker variability in speech processing*. San Diego: Academic Press. 145–165.

Kawahara, Shigeto (2015). The phonology of Japanese accent. In Haruo Kubozono (ed.) *The handbook of Japanese phonetics and phonology*. Berlin: de Gruyter Mouton. 445–492.

Keating, Patricia A. (1996). The phonology–phonetics interface. In Ursula Kleinhenz (ed.) *Interfaces in phonology*. Berlin: Akademie-Verlag. 262–278.

Kharlamov, Viktor (2012). *Incomplete neutralization and task effects in experimentally-elicited speech: evidence from the production and perception of word-final devoicing in Russian*. PhD thesis, University of Ottawa.

Kingston, John & Randy L. Diehl (1994). Phonetic knowledge. *Lg* **70**. 419–454.

Kinoshita, Keisuke, Dawn Behne & Takayuki Arai (2002). Duration and F0 as perceptual cues to Japanese vowel quantity. *Proceedings of the 7th International Conference on Spoken Language Processing* (*Interspeech 2002*). 757–760.

Kirby, James P. (2010). *Cue selection and category restructuring in sound change*. PhD dissertation, University of Chicago.

Klatt, Dennis H. & William E. Cooper (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. G. Nooteboom (eds.) *Structure and process in speech perception*. Berlin: Springer. 69–89.

Kluender, Keith R., Randy L. Diehl & Beverly A. Wright (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *JPh* **16**. 153–169.

Koenig, Laura L., Susanne Fuchs & Jorge C. Lucero (2011). Effects of consonant manner and vowel height on intraoral pressure and articulatory contact at voicing offset and onset for voiceless obstruents. *JASA* **129**. 3233–3244.

Kubozono, Haruo (1993). *The organization of Japanese prosody*. Tokyo: Kurosio.

Lacerda, Francisco (1995). The perceptual-magnet effect: an emergent consequence of exemplar-based phonetic memory. In Kjell Elenius & Peter Branderud (eds.) *Proceedings of the 13th International Congress of Phonetic Sciences*. Vol. 2. Stockholm: KTH & Stockholm University. 140–147.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum. 884–891.

Lindblom, Björn (1990). Explaining phonetic variation: a sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer. 403–439.

McCarthy, John J. (2007). *Hidden generalizations: phonological opacity in Optimality Theory*. Sheffield & Bristol, Conn.: Equinox.

McCarthy, John J. (2009). The P-Map in harmonic serialism. Ms, University of Massachusetts, Amherst. Available as ROA-1052 from the Rutgers Optimality Archive.

McCarthy, John J. & Alan Prince (1986). *Prosodic morphology*. Ms, University of Massachusetts, Amherst & Brandeis University.

McCarthy, John J. & Alan Prince (1993). *Prosodic morphology: constraint interaction and satisfaction*. Ms, University of Massachusetts, Amherst & Rutgers University. Available as ROA-482 from the Rutgers Optimality Archive.

McCawley, James D. (1968). *The phonological component of a grammar of Japanese*. The Hague & Paris: Mouton.

McCawley, James D. (1977). Accent in Japanese. In Larry M. Hyman (ed.) *Studies in stress and accent*. Los Angeles: Department of Linguistics, University of Southern California. 261–302.

Manaster Ramer, Alexis (1996a). Report on Alexis' dreams – bad as well as good. *JPh* **24**. 513–519.

Manaster Ramer, Alexis (1996b). A letter from an incompletely neutral phonologist. *JPh* **24**. 477–489.

Mańczak, Witold (1958). Tendences générales des changements analogiques. *Lingua* **7**. 298–325, 387–420.

Matsui, Mayuki (2015). *Roshia-go ni okeru yuusei-sei no tairitsu to tairitsu no jyakka: onkyo to chikaku*. [*Voicing contrast and contrast reduction in Russian: acoustics and perception*.] PhD thesis, Hiroshima University.

Mester, Armin (1990). Patterns of truncation. *LI* **21**. 478–485

Mitleb, Fares M. (1981a). *Segmental and non-segmental structure in phonetics: evidence from foreign accent*. PhD thesis, Indiana University.

Mitleb, Fares M. (1981b). Temporal correlates of 'voicing' and its neutralization in German. *Research in Phonetics* **2**. 173–191.

Moreton, Elliott (2004). Realization of the English postvocalic [voice] contrast in F_1 and F_2. *JPh* **32**. 1–33.

Mori, Yoko (2002). Lengthening of Japanese monomoraic nouns. *JPh* **30**. 689–708.

Nooteboom, S. G. & G. J. N. Doodeman (1980). Production and perception of vowel length in spoken sentences. *JASA* **67**. 276–287.

Nosofsky, Robert M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General* **115**. 39–57.

Oostendorp, Marc van (2008). Incomplete devoicing in formal phonology. *Lingua* **118**. 1362–1374.

Pater, Joe (2009). Weighted constraints in generative linguistics. *Cognitive Science* **33**. 999–1035.

Pierrehumbert, Janet B. (2001a). Exemplar dynamics: word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: Benjamins. 137–157.

Pierrehumbert, Janet B. (2001b). Stochastic phonology. *Glot International* **5**. 195–207.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology 7*. Berlin & New York: Mouton de Gruyter. 101–139.

Pierrehumbert, Janet B. & Mary E. Beckman (1988). *Japanese tone structure*. Cambridge, Mass.: MIT Press.

Piroth, Hans Georg & Peter M. Janker (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *JPh* **32**. 81–109.

Port, Robert F. (1996). The discreteness of phonetic elements and formal linguistics: response to A. Manaster Ramer. *JPh* **24**. 491–511.

Port, Robert F. & Penny Crawford (1989). Incomplete neutralization and pragmatics in German. *JPh* **17**. 257–282.

Port, Robert F., Fares Mitleb & Michael L. O'Dell (1981). Neutralization of obstruent voicing is incomplete. *JASA* **70**. S13.

Port, Robert F. & Michael L. O'Dell (1985). Neutralization of syllable-final voicing in German. *JPh* **13**. 455–471.

Poser, William J. (1984). *The phonetics and phonology of tone and intonation in Japanese*. PhD dissertation, MIT.

Poser, William J. (1990). Evidence for foot structure in Japanese. *Lg* **66**. 78–105.

Posner, Michael I. & Steven W. Keele (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology* **77**. 353–363.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.

Reed, Stephen K. (1972). Pattern recognition and categorization. *Cognitive Psychology* **3**. 382–407.

Rialland, Annie (1986). Schwa et syllabes en français. In Leo Wetzels & Engin Sezer (eds.) *Studies in compensatory lengthening*. Dordrecht: Foris. 187–226.
Rosch, Eleanor (1973). Natural categories. *Cognitive Psychology* **4**. 328–350.
Röttger, T. B., B. Winter, S. Grawunder, J. Kirby & M. Grice (2014). Assessing incomplete neutralization of final devoicing in German. *JPh* **43**. 11–25.
Scarborough, Rebecca (2003). Lexical confusability and degree of coarticulation. *BLS* **29**. 367–378.
Scarborough, Rebecca (2010). Lexical and contextual predictability: confluent effects on the production of vowels. In Fougeron *et al.* 557–586.
Slowiaczek, Louisa M. & Daniel A. Dinnsen (1985). On the neutralizing status of Polish word-final devoicing. *JPh* **13**. 325–341.
Slowiaczek, Louisa M. & Helena J. Szymanska (1989). Perception of word-final devoicing in Polish. *JPh* **17**. 205–212.
Smith, J. David & John Paul Minda (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**. 3–27.
Steriade, Donca (2000). Paradigm uniformity and the phonetics–phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.) *Papers in laboratory phonology V: acquisition and the lexicon*. Cambridge: Cambridge University Press. 313–334.
Steriade, Donca (2009). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: studies in honor of Paul Kiparsky*. Cambridge, Mass.: MIT Press. 151–179.
Sugito, Miyoko (1982). Tōkyō akusento ni okeru 'hana' to 'hana' no sēsē to chikaku. [The production and perception of 'flower' and 'nose' with a Tokyo accent.] In *Ni-hongo akusento no kenkyū*. [*Studies on Japanese accent*.] Tokyo: Sanseido. 182–201.
Syrett, Kristen & Shigeto Kawahara (2013). Production and perception of listener-oriented clear speech in child language. *Journal of Child Language* **41**. 1373–1389.
Taylor, Dennis Q. (1975). The inadequacy of bipolarity and distinctive features: the German 'voiced/voiceless' consonants. In Peter A. Reich (ed.) *The Second LACUS Forum*. Columbia, S.C.: Hornbeam. 107–119.
Thomas, Erik R. (2000). Spectral differences in /ai/ offsets conditioned by voicing of the following consonant. *JPh* **28**. 1–25.
Vance, Timothy J. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.
Vance, Timothy J. (1995). Final accent *vs*. no accent: utterance-final neutralization in Tokyo Japanese. *JPh* **23**. 487–499.
Warner, Natasha, Erin Good, Allard Jongman & Joan Sereno (2006). Orthographic vs. morphological incomplete neutralization effects. *JPh* **34**. 285–293.
Warner, Natasha, Allard Jongman, Joan Sereno & Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *JPh* **32**. 251–276.
Yu, Alan C. L. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology* **24**. 187–214.
Yu, Alan C. L. (2011). Contrast reduction. In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.) *The handbook of phonological theory*. 2nd edn. Oxford: Wiley-Blackwell. 291–318.
Zsiga, Elizabeth C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *JPh* **28**. 69–102.
Zue, Victor W. & Martha Laferriere (1979). Acoustical study of medial /t,d/ in American English. *JASA* **66**. 1039–1050.