

## Naturalness and phonetic grounding in Turkish vowel-height rules

Deepthi Gopal · Stephen Nichols

Received: date / Accepted: date

**Abstract** look at this super cool empirical data we found  
this tells us that bad phonetic cues + good abstract relationship = rare natural class  
so this article also fills a gap in both synchronic natural class typology and in the  
typology of the conditioning of sound change

**Keywords** phonology · sound change · Turkic languages

### 1 Introduction

**FIXME** Most classes are natural, but not all natural classes are equal; some seem to be much rarer than others. Trying to understand why things are this way should be caused (in whole or in part) by the trajectory of regular sound change, cite refs. and discuss EP vs PBP in a couple of lines. Implicit in dominant accounts of the conditioning of sound change (and the predictions about its typology) is the idea that the different environments in which a new phonological rule arises should constitute (equally?) good phonetic precursors for it. Fruehwald talks about cases in which we don't necessarily expect this to be the case; we argue here that not only is it not the case, but that even in cases which look like they have some representational unity ([sonorant]), not all precursors seem to work; the mismatch between poor cue strength and abstract representational relationship to the original triggering environment gives us both a cue to the early phonological conditioning of regular phonetic change, and a probabilistic prediction about the relative frequency of certain kinds of representationally-coherent class. Clue about phonological conditioning also lies in the fact that it's hard to 'stop at' sonorant — once you get there, you often extend further..

---

D. Gopal  
University of Cambridge  
E-mail: dg537@cam.ac.uk

S. Nichols  
University of Manchester

**FIXME further implicit in this argument is that we are talking about the end-products of sound change (blevins, RBO, frueh) and not the 'elaboration of' (read FR paper). if we're really talking about 'elaboration of', then we need some explicit theory of what that set can contain; argue that it is ambiguous whether we should see this representational control as 'end product' or 'further elaboration', but clear that it happens.**

look at this super cool empirical data we found  
this tells us that bad phonetic cues + good abstract relationship = rare natural class  
so this article also fills a gap in both synchronic natural class typology and in the typology of the conditioning of sound change

The major subsidiary goal of this paper is incremental empirical progress with respect to the phonology of modern Turkish. Although the cases of vowel harmony and of anomalous stress in Turkish are familiar to almost all phonologists, little attestation of the novel pattern we report here exists in the descriptive literature on Turkish, and none at all in the phonological literature. We argue that the current study fills a gap both in the state of language-specific knowledge, and more generally in the broad typology of the end-products of sound change.

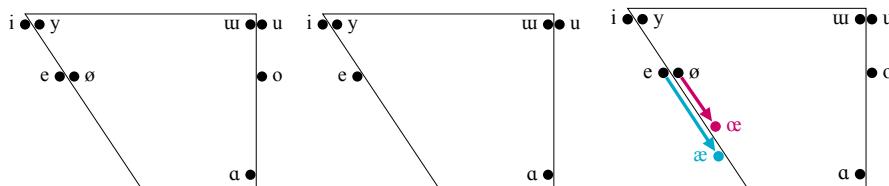
this paper is organised as follows. In section 2, we introduce the innovation in Turkish on which the rest of this paper focuses; section 5.2 discusses the similar but non-identical patterns that arise in non-standard varieties of Turkish, and section 2.3 provides the broader context of the typology of closed-syllable height effects. Section 3 contains a straightforward overview of the setup, methodology, and sampling of the study on which the remainder of the chapter is based. The acoustic data are addressed throughout section 4; section 4.1 deals with the overall distribution of results and the categoricity of the Turkish pattern, and section 4.2 considers the details of the behaviour of individual segments. We treat the various classes of exceptions to the alternation in section 4.4, and provide the overall perspective from the point of view of the hypothesised diachrony in section 5, along with some recapitulation of the state of inter-speaker variation and the typology in section 5.3.

## 2 Pre-sonorant mid vowel lowering in Turkish

In this section, we set out the case that the remainder of this study will investigate in detail; a note refreshing the reader's memory of well-known features of the Turkish phonological system is given in section 2.1, the pattern itself is described in section 2.2, and section 2.4 briefly motivates the experimental work whose setup and methodology appear in section 3. Detailed experimental results follow in section 4.

### 2.1 Notes on the overall system

For the reader's reference, we briefly contextualise the discussion with an overview of salient features of the Turkish phonological system. Turkish is generally understood in the phonological literature to have eight contrastive vowels (Hulst and Weijer 1991, Kabak 2011), whose phonological behaviour is well-known (for overview, Clements and Sezer 1982; Kabak 2011): processes of backness and rounding harmony both operate left-to-right across roots and suffixes, under the systematic constraint that rounding may only spread to high vowels in non-initial syllables. Although all eight vowels in the system are paired for backness and roundness, it is therefore the case that a gap arises in the non-initial syllables (figure 1) – the rounded mid vowels do not appear in these positions, save in loanwords. /e/ is therefore the only non-high front vowel that may appear in any position in the word, and the only vowel outside initial syllables that we can see as having 'room' to vary freely in height—/i/, /y/, /u/, /u/, and /ø/ are all constrained in one direction by the ceiling or the floor of the vowel space. The alternation in example 1 then looks like the right-hand diagram in figure 1: pre-sonorant realisations of /e/ and /ø/ *fill a gap* along the front diagonal of the vowel space.



**Fig. 1** Turkish vowels in initial syllables (left) and in non-initial syllables (center); sonorant-conditioned alternations shown in the Turkish vowel space (right).

The majority of the Turkic languages share similar systems of backness harmony, and most extend this to an allophonic distribution of *consonantal* segments (most strikingly in Karaim, in which the typical system of [back]-spreading appears to have been lost for the vowels, and retained instead as long-distance consonant harmony; Nevins and Vaux 2003). Clements and Sezer (1982) describe the Turkish velars /k/, /g/ and the lateral /l/ as being typically [+back] in the context of [+back] vowels and [-back] in the context of [-back] vowels, although subject to various blocking effects

and lexicalised exceptions. We consider (although not in exhaustive depth) the state of phonetic data on the backness of the lateral /l/ in section 4.2.2 below, with respect of the potential for degree of palatalisation to affect the lateral's suitability as a phonetic precursor to vowel-lowering; the experimental data conform well to the characterisation here.

The remainder of the sonorants are not, as far as the literature is aware, subject to alternations of this particular type, although the rhotic has variable realisation in general; it may be tapped, trilled, or approximant, and may be fricated, devoiced, or entirely deleted (the latter particularly in the imperfective /-iyor/ and the indefinite article /bir/). The literature is rather variable as to the state of the rhotic; although accounts generally coincide (Lewis, 1967; Göksel and Kerslake, 2005) in describing it as either an alveolar tap or trill, accounts of the extent to which it is subject to frication and devoicing vary. Lewis (1967, p. 7) and Kopkallı (1993, p. 29) claim that devoicing and frication are exclusively word-final, contradicting Blaskovics's (1964, p. 5–10) description of front-vowel-dependent frication in the rhotic. Comrie (1997) describes rhotic frication as exclusively *word-initial*; Yavuz and Balcı (2011, p. 25) claim that frication applies both word-initially and word-finally, but that devoicing is restricted to the word-final position. Due to the general constraints of experimentation, we make no strong claim here as to the status of the rhotic, other than the establishment of its broad tendency towards strongly devoiced realisation. Impressionistically, in our sample, all utterance-final rhotics were voiceless and fricative in realisation; intervocalic and pre-consonantal rhotics were generally flapped (although occasionally trilled, assimilated to an adjacent lateral, or approximant).

## 2.2 The case

In Turkish, the front mid vowels /e/ and /ø/ undergo alternations conditioned by the following coda, summarised ultimately (with reference to the results and descriptive discussion that follow in section 4) in examples 1, 3 and 4 below—/e/ and /ø/ are systematically low in a syllable closed by a sonorant, and /e/ is (for many speakers, but subject to significant inter-speaker variation; section 5.3) slightly raised relative to the non-sonorant mean in either pre-obstruent contexts, or in contexts in which the coarticulatory effect of high vowels is particularly significant. We focus in this paper particularly on the state of /e/, which is more frequent in the lexicon and displays more clearly categorical and phonetically discontinuous behaviour. Example 3 is illustrated in figure 2 with data from the study in section 4 for **one** representative (F09; female, b. 1980 Ankara) speaker. Pre-sonorant /e/ in un-affixed forms has F1 in the range 750–900 Hz; the corresponding alternant /e/ post-resyllabification has F1 400–600 Hz.

*Example 1* /e/ and /ø/ surface as [æ ~ a] and [œ] respectively in sonorant-closed syllables:

/erdem/	[ær.dæm]	‘virtue’
/hejkel/	[hej.kæl]	‘statue’
/gizem/	[gi.zæm]	‘mystery’

/biber/	[bi.bær]	'pepper'
/gøl/	[gœl]	'lake'
/gøm-mek/	[gœm.mek]	'bury-INF'

*Example 2* And, for some speakers, in syllables closed with /z/:

/pekmez/	[pek.mæz]	'treacle'
/merkez/	[mær.kez] > [mær.kæz]	'center'
/gel-mez/	[gæl.mæz]	'does not go'

*Example 3* Affixation destroys the environment for lowering (figure 2):

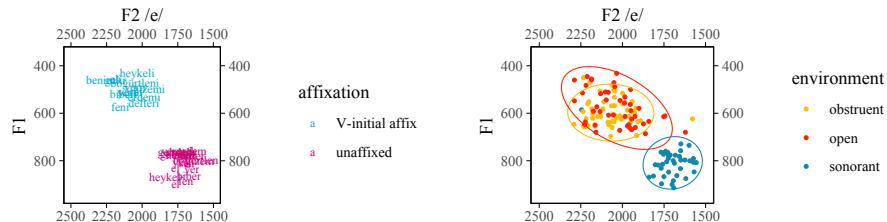
/erdem-i/	[ær.de.mi]	'virtue-ACC'
/hejkel-im/	[hej.ke.lim]	'statue-1SG.POSS'
/gizem-in/	[gi.pe.min]	'mystery-GEN'
/biber-in/	[bi.be.rin]	'pepper-GEN'
/gøl-i/	[gø.ly]	'lake-ACC'
/gøm-er/	[gø.mær]	'bury-3SG.P'

*Example 4* No lowering in other environments; some speakers show pre-obstruent /e/-raising:

/dede/	[de.de]	'grandfather'
/bebek/	[be.bek] ~ [be.bæk]	'baby'
/herkes/	[hær.kes] ~ [hær.kes]	'everybody'
/taze/	[ta.zə]	'fresh'

*Example 5* Unstressed open syllables preceding high vowels /i, y/: /e/ raised (section 4.3.1).

/deniz/	[di.niz]	'sea'
/kedi/	[kɪ.di]	'cat'
/be-nim/	[bɪ.nim]	'my'



**Fig. 2** F09: F1/F2 space (Hz) for alternating /e/. Let, pre-son. unaffixed *ben*, *biber*, *bögürten*, *defter*, *el*, *erdem*, *fen*, *gizem*, *heykel*, *sen*, *yer* & C-initial suffixed *vermek*, *gelmek* vs. pre-vocalic *benim*, *biberi*, *bögürteni*, *defteri*, *eli*, *erdem-i*, *feni*, *gelistir*, *gizemi*, *heykeli*, *senin*, *verir*, *yeri*. Right, all measurements coded by environment, 95% confidence ellipses.

We are not aware of any reference to this alternation in the formal phonological literature. Several general descriptions of the language are available, and vary in their characterisations of the system, in a manner consistent with the suggestion that the pattern under consideration is a recent development in Turkish. Lewis's (1967, 14) reference grammar describes raising in unstressed open syllables: 'a closer pronunciation, verging on the sound of **i**, especially in the first syllables of [...] **gece** 'night'', but mentions no lower allophone and no preconsonantal effects of any kind. Kornfilt (1997, 512), 30 years later, claims that an 'alternation phenomenon affects the front, non-high vowel [e] and [ø], which are lowered before sonorants in closed syllables'; she transcribes the lower allophone of [e] as [ɛ], although it's unclear to what extent this decision is impressionistic (or informed by real phonetic properties). Göksel and Kerslake (2005) give the distribution of /e/ as [æ] before sonorants, [ɛ] in stressed open syllables, and [e] elsewhere. We argue that none of these characterisations adequately represents the state of the system as we find it in current data, with the note that the disparity in descriptions corresponds to a significant time-separation and is concordant with further discussion of the relative recency of this pattern. In section 4, we present the production-study results that support the characterisation we propose in examples 1 and 3–5.

### 2.3 Coda-triggered height effects

The descriptive case(s) in Turkish that constitutes the majority of this paper can be placed at the intersection of two larger typologies of phonetically well-motivated phenomena: one of vowel quality effects conditioned by syllable structure, and the other of height effects triggered, irrespective of syllabic considerations, by the various sonorant segments. The pattern in Turkish either of these typologies: while closed-syllable vowel laxing is well-established, it is rarely predicated on the manner of articulation of the coda involved; and while rhotic-triggered height effects are common, they are rarely extended to the set of all sonorants and rarely dependent on syllable structure.

#### 2.3.1 Closed-syllable vowel laxing

It is reasonably well-established that there exists a general cross-linguistic tendency towards laxer vowels in closed syllables and tenser vowels in open syllables, both in distribution and in allophonic patterning; a straightforward example is the alternation that appears in French *nous gavottons* [nu ga.vo.tɔ̃] 'we gavotte' and *rigolo* [ʁi.go.lo] 'funny (masc.)', but *il gavotte* [il ga.vɔt] 'he gavottes' and *rigolote* [ʁi.go.lo.t̪] 'funny (fem.)'. Storme (2017c, p. 223–226) gives a non-exhaustive but substantial typological survey of 18 languages in which some such generalisation holds; for all these cases, 'lax' is understood acoustically to imply a two-dimensional movement in the vowel space, with a lax vowel having less peripheral F1 and F2 targets than a tense counterpart. Not all attested cases are equally general; in some languages laxing applies across the board to all vowels in closed syllables, while in others further restrictions on the affected vowel<sup>1</sup> or the consonantal context apply (example 6). It is unclear

<sup>1</sup> e.g. Nordhoff (2009): Sri Lanka Malay has the five-vowel inventory /a e œ œ u/, but only /œ œ/ alternate.

whether apparent statistical tendencies should be expected to hold more broadly or carry particular significance. If we were to take Storme's survey as broadly representative, we would observe that the mid vowels /e o/ are slightly under-targeted<sup>2</sup>, and that languages invoking segment-specific constraints do so particularly with respect to rhotics and dorsals<sup>3</sup>.

*Example 6* (Uma Juman) Kayan (Austronesian, Borneo): high vowels are lowered if followed by /h, l, r, ?/. (Blust, 2013, p. 263)

/laki?/	[lake?]	'male'
/uru?/	[uro?]	'grass'
/hivih/	['hi.veh]	'lower lip'
/duh/	[doh]	'female'
/bakul/	['ba.kol]	'basket'
/tumic/	['tu.mer]	'heel'
/?atur/	['?a.tor]	'arrange'

The existence of coda-conditioned effects on height and backness has been attributed in various cases (Féry, 2003; Botma and Van Oostendorp, 2012) to the existence of a close relationship between length, quality, and syllable structure; that is, contingent on the claim that vowels are shorter in closed syllables than in open syllables, which see e.g. Maddieson (1985). The duration-sensitive account of closed-syllable laxing then admits a connection to those patterns in which syllabically-conditioned length alternations arise, as example 7; but requires some form of the claim that longer vowels are tenser than shorter vowels.

*Example 7* Ingush (Nakh-Dagestanian, Ingushetia): underlying long vowels alternate with short vowels in closed syllables; short vowels don't alternate. (Nichols, 2011)

/duucə/	[du:cə]	'narrate-INF'
/duuc/	[duc:]	'narrate-PRS'
/niisə/	[ni:sə]	'straight'
/niis-lu/	[nis.lu]	'becomes straight'

One potential issue for explanations along these lines is that empirical generalisations about the relationship between quality and duration are variable, and difficult to straightforwardly align with the demands of articulation or the typology of laxing. The claim that closed-syllable laxing derives from the loss of duration in closed syllables requires that lax vowels be uniformly shorter than tense vowels; given a non-low vowel subject to laxing, we expect the lax counterpart to be lower than the tense one, and therefore must claim that higher vowels are longer than lower ones. However, it is in fact frequently the case across languages that the first formant correlates positively with duration, and that therefore height itself is inversely correlated with duration: that is, the lowest vowels (with the highest F1) are the longest vowels.

<sup>2</sup> Although available in all surveyed inventories but one, /e o/ undergo laxing in 7 of 17 languages; French (Southern), Chamorro, Indonesian, Kairiru, Palau, Sri Lanka Malay, Kuteb. The remaining languages surveyed target exclusively the more peripheral vowels, particularly /i u/.

<sup>3</sup> Patterns of the uneven distribution of vowel height in the context of uvulars and velars are well-known; for an overview, see Gallagher (2016, p. 103).

Claims to this effect appear both within-category and cross-category. If all other parameters are constant, Lehiste (1970) suggests that low vowels are longer than high vowels across the board, and posits physiological reasons relating to the gestural duration of jaw-opening; similar effects have also been attributed to the closeness of the jaw position during high vowels to that expected during most consonants (Maddieson, 1997; Gussenhoven, 2007). One of the more widely-known experimental results to this effect is Lindblom (1963)'s claim that the F1 of Swedish non-high vowels decreases (that is, vowel height increases) exponentially as vowel duration decreases. Similar correlations are reported in English (Peterson and Lehiste, 1960; Westbury and Keating, 1980), Hindi (Ohala and Ohala, 1992); and cf. Toivonen et al. (2015) for English and Swedish, and for a review of the literature. The relationship between duration and vowel height is revisited in sections 4.3 and 4.3.1. A portion of Storme (2017a,b,c)'s argument for a new account rooted in the perceptually-driven enhancement of post-vocalic contrasts between consonants is predicated on the claim, as above, that the derivation of lowering and centralising effects from the loss of duration is not justified; a full treatment of this account and its relationship to the case we consider here is beyond the scope of this paper.

The essential relationship of the Turkish case to the cross-linguistic pattern is largely straightforward, although in a few respects particularly divergent; the pattern in Turkish appears to show a larger deviation in phonetic space than the typical case of closed-syllable laxing, as should become apparent from the experimental data presented below, and the class of segments targeted is a new one. We argue that the pattern in Turkish is *lowering* rather than *lowering plus centralisation*, and that any apparent centralisation arises solely due to the inherent topology of the vowel space; unlike Storme (2017c, p. 95)'s case in French, for which F2 effects appear to be too large to be accounted for by height targets alone.

### 2.3.2 Sonorant-related height effects

Sonorant-triggered—particularly, rhotic-triggered—height effects are fairly well-attested cross-linguistically, both as non-contrastive differences in the contextual distribution of formant values in phonetic space, and as apparent phonotactics. What distinguishes the latter ‘phonological’ set from the patterns described in the previous section is the frequent absence of any restrictions relating to syllable structure; several cases of sonorant-conditioned laxing or lowering apply irrespective of the relative position of the sonorant trigger and the affected vowel. In the Austronesian language Thao (Blust, 2013, p. 264), all high vowels lower if adjacent to the alveolar flap /ɾ/ example 8; the requirement for adjacency applies in both tautosyllabic and heterosyllabic contexts, and the flap may precede or follow the affected vowel. (A further, more general case of laxing applies in the familiar set of closed-syllable environments.)

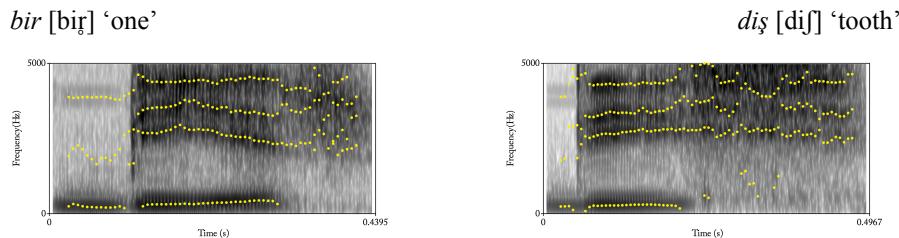
*Example 8* Thao (Austronesian; Taiwan): high vowels lower<sup>4</sup> adjacent to a rhotic. (Blust, 2013)

/rima/	[‘re.ma]	‘five’
--------	----------	--------

<sup>4</sup> Also if adjacent to /q/: /tuqrɪs/ [‘toq.res] ‘noose trap’, /qusaz/ [‘qo.sað] ‘rain’.

/rusaw/	[r̩o.saw]	'fish'
/ruʃ/	[r̩i.roʃ]	'saliva'
/lmir/	[l̩mer]	'grass, weeds'
/turu/	[t̩o.ro]	'three'
/hiburin/	[hiʃ̩.bo.ren]	'be mixed together'

In general, throughout the phonetically-informed literature, attestation exists of significant effects in F1 conditioned by the presence of an adjacent sonorant. These vary between (what we would diagnose as) the ‘phonetic’ and the ‘phonological’ in their scope, and in which sonorants they involve. What seems broadly uncontroversial is that strong articulatory and acoustic properties of the rhotics often favour the development of height effects in a pre-rhotic vowel. The most widely-cited acoustic characteristic of rhoticity as a whole is the lowered third formant (Ladefoged, 2003); but see Lindau (1985) for complications. Figure 3 shows spectra for a [ir]~[iʃ] pair in Turkish, for which there is a small but significant effect of the utterance-final rhotic on the preceding high vowel. The trill in particular (Recasens, 2002; Recasens and Pallarès, 1999; Solé, 2002) has been shown to force tongue dorsum lowering and retraction, which Proctor (2009) suggests (for Latin American Spanish) is a more general property of the whole class of coronal liquids. Bradley (2010), in a survey of Ibero-Romance, claims—unlike our case—that the effect of the post-vocalic trill on the height of the preceding vowel is strongest in the \_rV context, and weaker in \_rC, and attributes this to constraints on articulatory organisation.



**Fig. 3** For one speaker (F01, Istanbul/1997), spectra with formants indicated for *bir* [bir̩] ‘one’ and *diʃ* [diʃ] ‘tooth’, showing formant movement driven by the presence of the rhotic.<sup>6</sup>

Mid vowel lowering before a rhotic coda is attested widely: throughout Ibero-Romance (Bradley, 2010), and in the context of the French *loi de position* (Storme, 2017b); in Swedish for /ɛ/ and /ø/ (Riad, 2014), in Faroese /e/ (Árnason, 1999), and, as before, in the examples for Schaffhausen German that we give in section 5.1.1. A detailed discussion of the situation of the laterals appears in section 4.2.2, and we reserve a fuller overview for those passages. We draw particular attention to accounts of disparities between the rhotics and the laterals in degree and even direction of F1-effect; West (1999) gives, for southern British English, significantly higher F1 in rhotic contexts than in lateral ones, and attendantly the prediction that lateral-triggered lowering effects should, in phonetically-controlled cases, be smaller; but note that this lateral is

<sup>6</sup> This is an utterance-final token; the rhotic is quite fricated.

necessarily, due to the well-known state of English phonotactics, a velarised or ‘dark’ one. Phonetically non-velarised laterals are often ignored by vowel-lowering rules, even after further generalisations and extensions (as with the case for Schaffhausen German given in section 5.1.1); we therefore consider the state of the Turkish lateral in section 4.2.2.

In European Portuguese, Vigário (2002, p. 86–88) describes a fairly similar rule; the non-high vowels /e o/ must lower in word-final, unstressed syllables closed by a sonorant<sup>7</sup> (the rhotic, the (velarised) lateral, or the nasals/nasal vowels), neutralising the contrast in /e o/–/ɛ ɔ/ and overriding the output of a general process that forces the raising of low vowels pre-nasal. Thus<sup>8</sup>, *revólver* [ɾi.ˈvɔl.vɐɾ] ‘revolver’, *júnior* [ʒu.ni.ɔɾ] ‘junior’, *nível* [ni.veɫ] ‘level’, *álcool* [aɫ.kɔɫ] ‘alcohol’, *sémen* [se.men] \*[se.min] ‘semen’, *cólofon* [kɔ̃.lu.fun] \*[kɔ̃.lu.fun] ‘colophon’. This reflects both a nice case of a similar alternation, and the (possible) phonologisation of the variable correlates of the nasals: although anticipatory nasalisation should drive an increase in F1 (Krakow et al., 1988), the introduction of the ‘nasal formant’ (Beddor, 1993; Beddor et al., 1986) drives a reduction in the perceptual space of the nasalised vowels, causing perceptual *raising* in low-mid and low vowels. Where the Portuguese case differs from the Turkish one is in the systematic relationship to stress and plausibly therefore to duration, which we will revisit in section 4.3.

## 2.4 This study

In the preceding sections, we have addressed the extent to which we can coerce the Turkish pattern into a broader typology of patterns with seemingly similar phonetic and phonological conditioning. With the set of potential phonetic precursors in mind, we present below the results of our experimental investigation of the status of pre-consonantal height effects in the (standard) Turkish vowels, the evidence for their categoricity, and the details of their scope and conditioning.

One of the aims of this investigation is necessarily descriptive. We are not aware of a previous systematic investigation of height effects in Turkish vowels, or any formal analysis thereof; existing descriptions combine impressionistic auditory judgments and introspection, and as such require experimental confirmation. The experimentation described in the following section is a first approach to the descriptive problem; our discussion of the results in section 4 is framed around the issues of cross-linguistic typology and the phonetic precursors to phonologisation that we have raised in this section and in the introduction respectively.

## 3 Methodology and sample

All analysis that follows reflects data from a production study conducted between January 2016 and June 2017 in Manchester. Thirteen native speakers of Turkish, all

<sup>7</sup> Not in non-final syllables: *revertér* [ɾi.vir.ˈter] \*[ri.ver.ˈter]; except for /l/, which can trigger neutralisation across the board: *delgado* [dɛɫ.ˈga.du] ‘thin’, *relvinha* [ʁɛɫ.ˈvi.nɐ] ‘grass.dim’.

<sup>8</sup> Vigário (2002) gives the relevant vowel only; all errors in the remainder of the IPA rendering here are mine.

resident at the time of experimentation in the UK, read a list of 220 items, along with a further list of 20 sentences containing target /e/ embedded in varied phonological and morphological environments. Speakers' ages ranged from 20-39; there were eleven female speakers and three male. All speakers were raised in the place of origin within Turkey (as listed in table 1); length of time resident outside Turkish-speaking areas ranged from 0.5 to 8 years. Speakers were not made aware of the purpose of the experiment prior to recording, and participation was voluntary—no payment was offered.

The preponderance of female speakers was an unfortunate artefact of the process of speaker recruitment. As such, more data is required to determine whether sex has a statistically meaningful predictive effect on the nature of the pattern, and we cannot comment further; we exclude data points corresponding to male participants from the overall statistical analysis and modeling, but these data are included in qualitative observations where appropriate.

ID	Year of birth	Place of origin	ID	Year of birth	Place of origin
F01	1997	Istanbul	F08	1982	Ankara
F02	1995	Istanbul	F09	1981	Istanbul
F03	1991	Istanbul	F10	1980	Ankara
F04	1988	Izmir	F11	1978	Ankara
F05	1987	Istanbul	M01	1989	Kayseri
F06	1985	Fethiye	M02	1985	Denizli
F07	1983	Bursa	M03 <sup>9</sup>	1980	Kars

**Table 1** Speaker metadata (index, year of birth, region of origin) for all participants.



**Fig. 4** Reference map of participants' origins<sup>10</sup>.

120 instances of /e/ appeared in the set of test items, in obstruent-closed (42), sonorant-closed (40), and open (38) syllables. 70 total were (primary-) stressed and

<sup>9</sup> Represents a fairly clear instance of dialectal divergence and excluded from overall calculations; see figure 37 for discussion.

<sup>10</sup> R: ggmap (Kahle and Wickham, 2013), ggalt (Rudis et al., 2017).

the remainder unstressed<sup>11</sup>. The test set contained 32 instances of /ø/ (8 pre-obstruent and open; 16 pre-sonorant). Due to distributional restrictions on Turkish vowels, /ø/ almost never appears outside initial syllables, and is lower-frequency than /e/. The distribution of /ø/-containing tokens was therefore necessarily particularly skewed with respect to stress—the pre-obstruent and pre-sonorant categories were evenly split, but all but one of the /ø/ in open syllables were unstressed: the French loan *banliyö* ‘banlieue’ was included, but not all speakers found it acceptable. The remaining items provided data on the non-target vowels. A full list of items tested is provided in the appendix, in the standard Turkish orthography with phonemic transcription and gloss; those expected to be exceptional have been marked.

environment	e	ø	control
no coda	i.'le 'with'	bø.'lym 'chapter'	da.ki.ka 'minute'
coda obstruent	fj.ka.jet 'complaint'	'tʃøp 'garbage'	has.ta 'ill'
coda sonorant	bi.'ber 'pepper'	ɔr.'nek 'example'	sa.'mur 'sable'

**Table 2** Example items

For each speaker, one repetition of the wordlist and sentence-list was recorded. Recordings were made in a quiet room using a Sony PCM-M10 linear PCM recorder and an Audio-Technica ATR1200 cardioid microphone, and saved sampled at 44.1 kHz. Speakers were encouraged to read at a comfortable pace, and to correct themselves if unhappy with their own production. (120 stimuli + 20 sentence-internal targets)\*(13 speakers) = 1820 total instances of /e/ and 416 total instances of /ø/ were recorded, minus 13 /ø/ and 14 /e/ that occurred in lexical items that individual speakers found unacceptable. 55 /e/ tokens (3%) were removed due to the diagnosis of systematic and conditioned exceptionality, and treated independently. 10 /e/ (0.55%) and 19 /ø/ (4.6%) of these were excluded due to deletion or devoicing, mispronunciation/production errors, segmentation difficulties, interference from non-modal voicing, and post-palatalisation coarticulation, leaving 1746 instances of /e/ and 383 instances of /ø/ for analysis. 2511 tokens were measured for the remaining 6 underlying vowels as comparison; since these were non-target vowels, no attempt was made to balance their distribution (/a/ 843, /u/ 258, /o/ 234, /i/ 560, /y/ 366, /ɯ/ 250).

A further set of exploratory data were collected to test systematic patterns of exceptionality, representing a further 300 tokens of /e/. These were excluded from the overall statistical analysis, and were not distributed evenly among speakers: the state of this extra dataset is discussed in detail in subsection 4.4.

All segmentation and acoustic analysis were carried out in Praat (Boersma and Weenink, 2017); target vowels were extracted from the produced tokens, with boundaries inserted manually based on visual inspection of the waveform and spectrogram. In pre-/post- obstruent and pre-pausal contexts, the beginning and end of the target

<sup>11</sup> In practice, what this means is that these target vowels were in final and non-final syllables respectively: Turkish stress is typically word-final when regular, and although exceptionally-stressing items exist (on which Inkelas 1999; Kabak and Vogel 2001; Inkelas and Orgun 2003, *inter alia*) these were not tested.

vowel was established with a reasonable degree of confidence by the presence of low-frequency periodic noise and clearly-visible formant structure. For vowels post-sonorant, the initial boundary was placed at the onset of the formant steady-state. Pre-nasal, the appearance of the high-amplitude formants associated with a nasal release was abrupt enough to provide a high-confidence end boundary, but for laterals and rhotics the cues to the transition (formant values, intensity) are typically more gradual; pre-pausal /r/ showed significant frication, distinguishing it unambiguously from the target, and the final boundary pre-lateral was placed where a significant change in formant trajectory and intensity coincided.

F1 and F2 (in Hertz) were measured at 3 points (25%, 50%, and 75% of inter-boundary duration) and averaged, and duration (in milliseconds) was automatically exported by a Praat script. All statistical analysis was done in R (R Core Team, 2017). Formant data were then normalised to formant-extrinsic z-scores (Lobanov-normalised; Lobanov 1971; Adank et al. 2004) for comparison across speakers, using the R `phonR` package (McCloy, 2016).

## 4 Data

In this section, we present the acoustic data relating to coda-conditioned height effects in Turkish. This begins (section 4.1) with a demonstration of the patterns' robustness and apparent phonological categoricity; section 4.3 deals with the patterning and condition of vowel duration. Details of lexical exceptions and blocking environments appear in section 4.4, individual triggering segments in section 4.2, and the (incipient) generalisation to some voiced obstruents is considered in section 4.5.

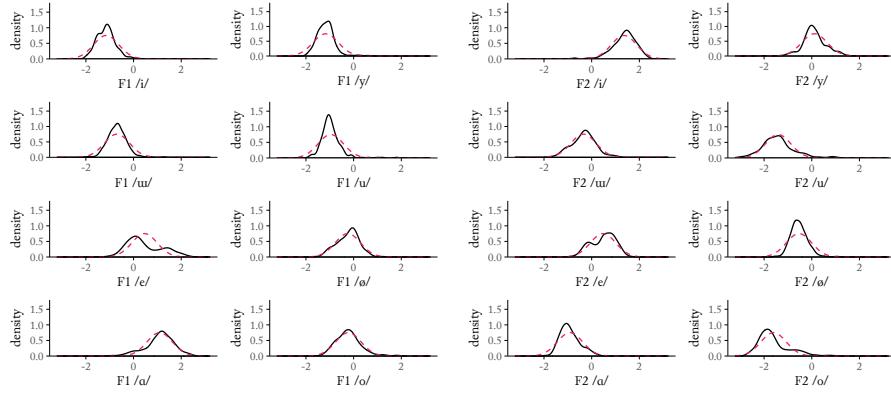
Since the focus in this discussion is on the comparative state of the system rather than specific physical-acoustic properties for which a rescaling to a Hertz-like value might be of interest, we have presented the Lobanov-normalised/z-score measure directly, throughout. For /e/, the range of non-normalised F1 values is [346.98, 1178.94] (Hz), and of F2 values [1136.88, 2828.44]; for /ø/, the F1 range is [334.02, 851.42], and the F2 range is [1062.01, 2329.58].

### 4.1 Distribution and categoricity

In figure 2, we presented a sample of evidence that pre-sonorant /e/ in forms like unaffixed *heykel* and affixed *heykeli* shows distinct realisations. In this subsection, we demonstrate that these distinct, and distributionally discontinuous /e/-realisations correspond strictly to the nature of the following coda, and claim that this is indicative of a categorical, phonological effect targeting /e/ in syllables closed by a sonorant coda.

How do we diagnose a *phonologically*-conditioned effect? The classical criterion that we implicitly cite in example 3 and figure 2 is the existence of morphologically-sensitive *alternation*; affixation destroys the environment for lowering. One straightforward approach when faced with unstructured empirical data is to search for bimodality in the acoustic signal: if some phonetic parameter P has a discontinuous

distribution, the probability that P is subject to some phonological effect increases<sup>12</sup>. Probability density functions<sup>13</sup>(PDFs) on (Lobanov-normalised) F1 and F2 are given in figure 5, representing the distribution of all tokens of each ‘phonemic’ vowel category in the experimental result; an idealised Gaussian distribution with identical mean and standard deviation has been overlaid to aid the reader’s eye in assessing deviation from a normal model. F1 and F2 distributions for /e/ are strongly bimodal (i.e. have two local maxima); Hartigan’s dip test indicates that the distribution of both parameters is not unimodal (strongly so for F1:  $p = 0.03597$ ; weakly for F2:  $p = 0.08897$ ).



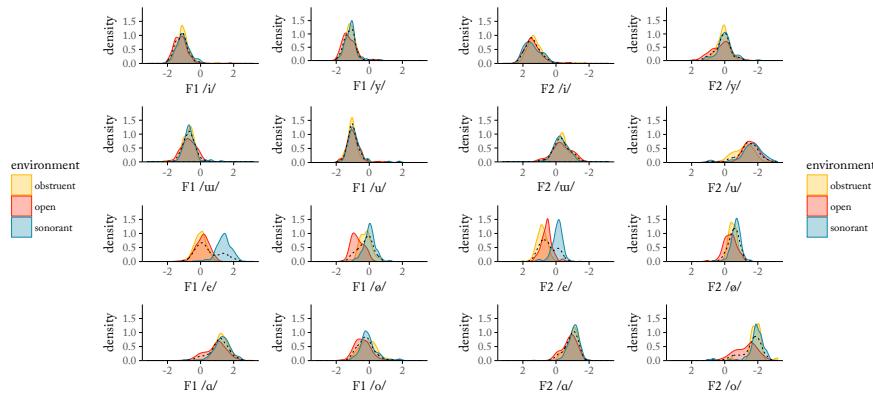
**Fig. 5** Probability density functions for F1 and F2, Lobanov-normalised across all speakers, in 8 vowel categories, with overlaid idealised Gaussian.

Figure 6 suggests that this bimodality derives from the composition of distinct distributions for the contextual realisations of /e/. If the sample is decomposed by context (syllable closed by *obstruent*, syllable closed by *sonorant*, open syllable), basis density functions can be calculated post-resampling whose maxima differ, but whose individual behaviour is no longer multimodal; the modes for these recalculated functions correspond to the multiple maxima in figure 5. The probability density function across all tokens for /ø/ shows no obvious bimodality on initial examination, but an apparent effect arises—especially in F1—on visual inspection when decomposed by context<sup>14</sup>.

<sup>12</sup> For complications, cf. Scobbie (2005): broadly, phonetic discontinuity signals categoricity, continuous distribution signals gradience; but this is subject to various confounds and empirical difficulties.

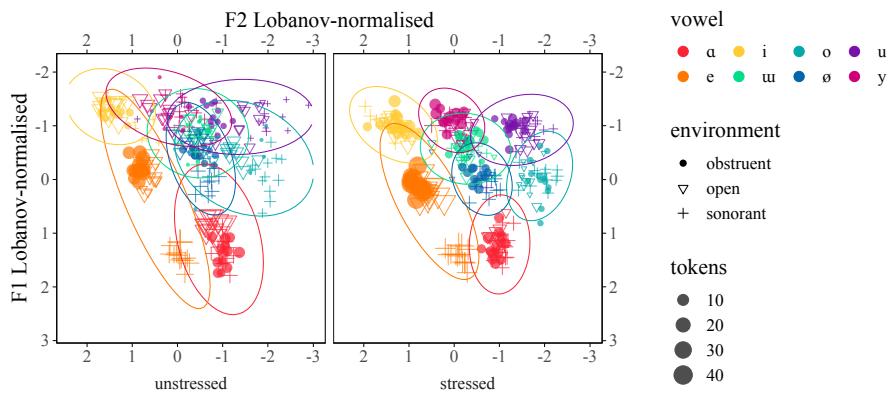
<sup>13</sup> These and further PDFs following are calculated using ‘kernel density estimation’ (Silverman, 1986), which *non-parametrically* estimates the probability density function for a random variable given a finitely-sized data sample: being unbound to any predetermined shape, kernel density estimates are better able to capture the presence or absence of irregularities in the distribution of data than parametric estimates. In base R: function `density`; in `ggplot2`: `stat_density`.

<sup>14</sup> The existence of multiple ‘real’ modes in a sample can be masked by insufficient overall distance, as with the superposition of two normal distributions whose means differ by less than the sum of their standard deviations—see for overview Schilling et al. 2002



**Fig. 6** F1 & F2 probability density functions, by environment. Dotted line: overall PDF from figure 5.

#### 4.1.1 The overall picture

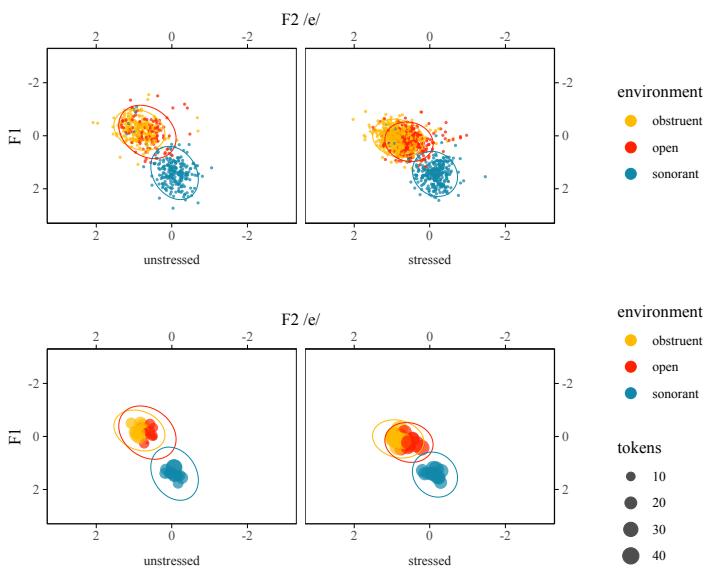


**Fig. 7** F2×F1 space, by vowel category and following environment: each point represents the mean measurement for a single speaker's productions of a particular vowel in the indicated environment; the size of a point denotes the number of tokens represented. 95% confidence ellipses indicate the overall spread of the *token-by-token* data for a particular vowel.

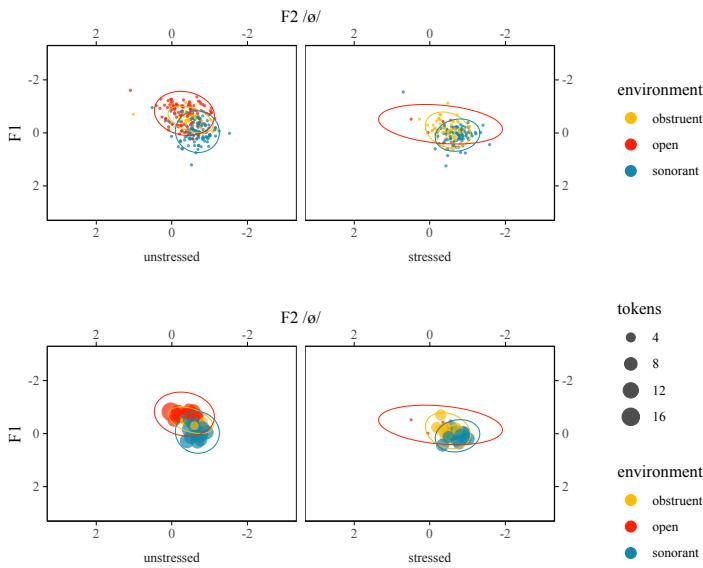
Figure 7 shows the Lobanov-normalised F2×F1 space for all 11 *female* speakers, averaged over all tokens across the 8 underlying vowel categories and separated by stress<sup>15</sup>. Each point represents the mean value for a single speaker's productions of a given vowel in a given context, with point size scaled by the number of tokens considered. In figures 8 and 9, F2×F1 space for /e/ and /ø/ are shown individually, both token-by-token and averaged over the speakers in the sample.

<sup>15</sup> For all figures in this paper, ‘environment’ refers to the following segment, and ‘unstressed’ and ‘stressed’ may safely also be read as ‘non-final’ and ‘final’ respectively.

The clear initial generalisation that we reaffirm here is that realisations of /e/ in pre-sonorant contexts diverge strongly and unambiguously from those in other contexts, subject to very little variation; realisations of /ø/ in pre-sonorant contexts also diverge from those in other environments, but the effect appears to be much smaller, and we might claim that there is more variation between speakers. One further striking observation that emerges naturally from figure 7 is the prevalence of strong height effects, presumably coarticulatorily-driven, for open vowels in non-final (that is, unstressed) syllables; the tendency of vowels in unstressed open syllables to be higher (have lower F1) is weak only for /e/.



**Fig. 8** F2×F1 space for /e/, by following environment. Top: each point corresponds to a single token. Bottom: each point corresponds to the mean measurement for a single speaker's productions, with size corresponding to number of tokens. Ellipses indicate regions of 95% confidence for the token-by-token data in both plots.

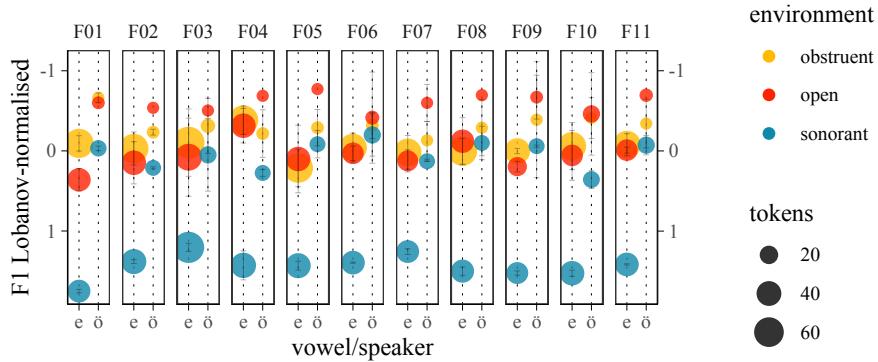


**Fig. 9**  $F2 \times F1$  space for  $/ø/$ , by following environment. Top: each point corresponds to a single token. Bottom: each point corresponds to the mean measurement for a single speaker's productions, with size corresponding to number of tokens. Ellipses indicate regions of 95% confidence for the token-by-token data in both plots.

Several preliminary generalisations emerge from figures 7–9, and from the preceding overview of density. In  $F2 \times F1$  space, tokens corresponding to pre-sonorant  $/e/$  have essentially no overlap with those corresponding to  $/e/$  in an unclosed or obstruent-closed syllable; the majority of the apparent effect is in  $F1$ , with a smaller apparent effect in  $F2$  corresponding roughly to movement along the front diagonal of the vowel space, which we may ascribe entirely to height rather than backness. The relative relationships of the pre-obstruent and open-syllable realisations of  $/e/$  are ambiguous and appear to have some dependence on stress; the overlap is near-complete for unstressed tokens, but stressed pre-obstruent  $/e/$  appears to correspond to slightly increased  $F2$  and slightly reduced  $F1$ . The data for  $/ø/$  indicate that some relationship exists between environment and  $F1$ ; the main contrasts relative to the state of  $/e/$  lie in the lack of distributional discontinuity (although pre-sonorant realisations have the largest  $F1$  values, there is overlap between their range and the ranges for the other environments), and in ordering: for  $/ø/$ , realisations in unstressed *open* syllables are consistently the most close (have the lowest value for  $F1$ ).

Although figures 8 and 9 include speaker averages, they do not indicate the correspondence between individual points and speakers *across categories*. In figure 10, averaged  $F1$  measurement is visualised, by speaker and environment, for both the mid vowels  $/e\ ø/$ ; this is more indicative of the *variation* in the relationship between the three major contexts (and in particular, the relationship between pre-obstruent mean and the open-syllable mean) by individual. Pre-sonorant  $/e/$  is invariant across speakers in its separation from other contexts; the major locus of differences between speak-

ers is the relationship between pre-obstruent /e/ and open-syllable /e/, which varies to the extent of undergoing reversals between individual speakers. Pre-sonorant /ø/ also follows a general speaker-independent trend, but again, individuals vary in the extent to which category separation arises for the other environments. It is reasonable to ask here whether the various loci of individual variation are substantially cross-correlated: do certain /e/-systems correspond to particular /ø/-systems? We revisit this question in section 5.3, and throughout the discussion that follows, we discuss and model the extent to which the effects that we claim here can be held to be robust and significant.



**Fig. 10** Normalised F1 by individual speaker, for /e/ and /ø/, averaged over stressed and unstressed contexts. Error bars show one standard deviation.

#### 4.1.2 Significance and effect size

The previous sub-section offers a broad overview of the environment-mediated patterns visible in the experimental dataset. In the discussion that follows, our aim is to explore some of the available detail; to begin with, we consider the statistical relationship of vowel height to the various possible predictors. In service of this, we fit standard linear mixed-effects models to subsets of the data for /e/ and /ø/ (using R: lme4, Bates et al. 2015), with F1 as the dependent variable, and the following environment ('obstruent', 'sonorant', or 'open'), stress ('stressed' or 'unstressed'), and their interaction as categorical fixed effects. Potential contributions from individual speaker and lexical item were treated as random effects: models include a random slope of environment and stress by speaker, and of stress<sup>16</sup> by word. (Another linear mixed-effects model was fitted to investigate the effect of individual sonorant segments (m, n, l, r); for the presentation and discussion of these latter results, see section 4.2.) The discussion in following sections, unless otherwise specified, drops the male speakers from the sample.

The choice of mixed-effects models is informed by the possibility of idiosyncracy. There is no a priori reason to expect that individual speakers respond identically to environment or stress; nor is there any reason to expect that all lexical items are treated

<sup>16</sup> Not for /ø/, due to the unbalanced nature of the test set; instances of stressed /ø/ were so infrequent that models incorporating a random slope here *do not converge*.

unexceptionally. The incorporation of random effects into a mixed model is then designed to account for this; the inclusion of the random slope of stress by speaker, for example, allows the effect of stress (on vowel height) to vary across speakers (different speakers may well implement cues to stress differently, colouring the resultant production).

Summaries of these models are presented below in tables 3–6. These are presented in a standard format<sup>17</sup>; as a brief guide, these should be interpreted by the reader as follows. The ‘base’ levels assumed by the model are ‘obstruent’ and ‘unstressed’: the estimate for the mean (*z*-score normalised) F1 in unstressed, pre-obstruent contexts is then given by the intercept, –0.2074. In tables corresponding to fixed effects, estimates for further terms predict the difference between these levels and the relevant ‘base’ ones: thus from table 3, the predicted difference between {open, unstressed} and {obstruent, unstressed} falls in the confidence interval [–0.1550, 0.1330] (indistinguishable from the null, or zero-effect, hypothesis), and the predicted difference between {open, stressed} and {open, unstressed} falls in the interval [0.0077, 0.2663] (F1 is expected to increase with stress in open syllables).

In tables corresponding to random effects, ‘variance’ and ‘SD’ (standard deviation) are, as in their usual definition, to be interpreted as measures of the spread corresponding to each parameter; that is, in this case, indices of the expected difference in the final value of F1 caused by idiosyncratic speaker or word effects. Correlations represent the extent to which a given fixed effect is invariant for a given random variable. As an example, the perfect correlation between *word* and *stress* in table 4 really just tells us that given a particular word, ‘stress’ does not vary (that is: /e/ in a particular word is always either stressed or unstressed), and as such is entirely unsurprising. In some situations, a correlation of this type indicates a model that is too complex for the data; however, in this case, there is a theoretical justification for the retention of the stress-word interaction. Even though the value of stress will always remain the same given a particular vowel in a word, there is no reason to expect the *phonetic* effect of that stress on vowel height to be invariant across words; retaining the term in the model allows us to comment on the extent to which the relationship between F1 and stress changes across lexical items.

As an index and as an analytically sound alternative to ‘*p*-values’ in a case in which these are non-meaningful<sup>18</sup>, 95% confidence intervals are given for each fixed term in the model, calculated via semi-parametric bootstrap estimation. Parameters whose value is statistically significant by this measure (that is, non-inclusive of the null) have been marked \*. As a guideline, *t*-values greater than 2 should be considered significant.

<sup>17</sup> We’re aesthetically indebted here to Fruehwald (2016) for the decision to tabulate density plots for the bootstrapped parameter estimates alongside the model results, and we have done the same throughout.

<sup>18</sup> Cf. Bates et al. (2015): in complex models involving e.g. partially-crossed designs, the null distributions of parameter estimates are not *t*-distributed for finite samples, but computing a *p*-value based on the degrees of freedom for a *t*-distribution (or symmetrically, estimates based on the *F* statistic) makes such asymptotic assumptions. The well-supported alternative, for the purpose of deciding on significance, is bootstrapping—simulating data from the fitted model, refitting the model, and re-extracting new estimated parameters—from which confidence intervals can be calculated; if the null hypothesis could correspond to points within the confidence interval for a given parameter (i. e. the parameter’s effect could be zero), then it cannot be rejected in that context.

Parameter	Estimate	95% confidence	Bootstrap	<i>t</i> -value
Intercept: obstruent	-0.21436	[-0.3086, -0.1086]		-4.267
Environment: open	-0.03109	[-0.1550, 0.1330]		-0.505
Environment: sonorant*	1.61723	[1.510, 1.837]		18.716
Stress: stressed*	0.30500	[0.1938, 0.3791]		8.435
open × stressed*	0.15163	[0.0077, 0.2663]		2.464
sonorant × stressed*	-0.37112	[-0.5173, -0.2404]		-5.205

**Table 3** Estimated means and derived confidence intervals for fixed effects in the model  $F1 \sim \text{environment} \times \text{stress} + (1 + \text{environment} + \text{stress}|\text{speaker}) + (1 + \text{stress}|\text{word})$  for /e/. 95% confidence intervals are calculated on 6,000 semi-parametric bootstrap estimates (R: `bootMer` from `lme4`), for which density plots are shown in the final column.

Group		Variance	SD	Correlation	
<i>Word</i>	Intercept	0.119126	0.34515		
	Stress: stressed	0.150095	0.38742	-1.00	
<i>Speaker</i>	Intercept: obstruent	0.010646	0.10318		
	Environment: open	0.021631	0.14707	-0.18	
	Environment: sonorant	0.036538	0.19115	-0.52	0.51
	Stress: yes	0.007788	0.08825	0.33	-0.23
<i>Residual</i>		0.105680	0.32508		

**Table 4** Random effect standard deviations and correlations, model  $F1 \sim \text{environment} \times \text{stress} + (1 + \text{environment} + \text{stress}|\text{speaker}) + (1 + \text{stress}|\text{word})$  (table 3), /e/.

It's evident from tables 3 and 5 that the predicted difference between tokens in sonorant-coda *unstressed* syllables and all other unstressed tokens is large and statistically significant for both /e/ and /ø/, as expected; the interaction with stress is substantial. The presence of (primary) stress reduces F1 in pre-sonorant /e/, but does not remove the large context separation; since, however, this is difficult to clearly adduce from the coefficients presented, as not every possible ordered pair of two-factor interactions is compared by the model<sup>19</sup>, table 7 gives the relevant terms and estimates of significance from another mixed-effects model fitted for the *stressed* subset of /e/-tokens alone, with random effects retained as above and the single fixed effect of environment. Modeling predicts no difference between pre-obstruent /e/ and /e/ in open syllables if both are unstressed; table 7 confirms that in stressed contexts, these differ, with  $F1_{\text{obstruent}} < F1_{\text{open}}$ . We argue that separating out the stressed and unstressed contexts here is not an unprincipled decision; there is a reasonably robust cross-linguistic basis for the claim that these might differ in the phonology, and there *is* an overall significant effect of stress itself on vowel quality. In sections 4.3 and 4.3.1,

<sup>19</sup> That is: no coefficient directly compares e. g. {open, stressed} and {sonorant, stressed}

we argue that there is indeed a larger interaction with stress-conditioned height and duration effects.

Parameter	Estimate	95% confidence	Bootstrap	t-value
Intercept: obstruent*	-0.4835	[-0.7332, -0.2475]		-3.410
Environment: open*	-0.3351	[-0.6154, -0.0483]		-2.482
Environment: sonorant*	0.3986	[0.1083, 0.6958]		2.479
Stress: stressed*	0.3201	[0.0419, 0.6169]		2.140
open × stressed <sup>20</sup>	0.2117	[-0.3531, 0.7609]		0.822
sonorant × stressed	-0.1411	[-0.4973, 0.2118]		-0.872

**Table 5** Estimated means and derived confidence intervals for fixed effects in the model  $F1 \sim \text{environment} \times \text{stress} + (1 + \text{environment} + \text{stress}|\text{speaker}) + (1|\text{word})$  for /ø/. 95% confidence intervals are calculated on 6,000 semi-parametric bootstrap estimates, for which density plots are shown in the final column.

Group		Variance	SD	Correlation	
<i>Word</i>	Intercept	0.04007	0.2002		
<i>Speaker</i>	Intercept: obstruent	0.01806	0.1344		
	Environment: open	0.02618	0.1618	-0.95	
	Environment: sonorant	0.02657	0.1630	-0.16	0.46
	Stress: yes	0.02133	0.1461	-0.49	0.29
<i>Residual</i>		0.07373	0.2715		

**Table 6** Random effect standard deviations and correlations, model  $F1 \sim \text{environment} \times \text{stress} + (1 + \text{environment} + \text{stress}|\text{speaker}) + (1|\text{word})$  (table 5), /ø/.

Parameter	Estimate	95% confidence	Bootstrap	t-value
Intercept: obstruent	0.08364	[-0.0154, 0.1801]		1.663
Environment: open*	0.11841	[0.0093, 0.2248]		2.147
Environment: sonorant*	1.32594	[1.183, 1.472]		17.674

**Table 7** Estimated means and derived confidence intervals for fixed effects in the model  $F1 \sim \text{environment} + (1 + \text{environment}|\text{speaker}) + (1|\text{word})$  for stressed /e/ only. 95% confidence intervals are calculated on 6,000 semi-parametric bootstrap estimates (R: `bootMer` from `lme4`), for which density plots are shown in the final column.

<sup>20</sup> Recall that only four data points went into this estimate.

Group		Variance	SD	Correlation
<i>Word</i>	Intercept	0.02551	0.1597	
<i>Speaker</i>	Intercept: obstruent	0.02285	0.1512	
	Environment: open	0.01548	0.1244	-0.20
	Environment: sonorant	0.04737	0.2176	-0.83 0.40
<i>Residual</i>		0.09305	0.3050	

**Table 8** Random effect standard deviations and correlations, model  $F1 \sim \text{environment} \times \text{stress}$   
+ (1 + environment + stress | speaker) + (1 | word) (table 7), stressed /e/.

There is a real, persistent, and statistically-significant effect of environment on the first formant for both /e/ and /ø/. Our inability to diagnose a statistically meaningful split between pre-obstruent and open-syllable /e/ we take as *inconclusive* rather than definitively *negative*, and revisit the problem in the discussion of individual variation in section 5.3. Effects are given in *z*-score measures throughout, which scale the whole dataset to a normal distribution with mean 0 and standard deviation 1; in order to grasp the meaning of the expected values given here, each linear-model estimate can then be understood to correspond to a number of standard deviations across the whole vowel space. This is not small; as an index of comparison, one standard deviation in F1 across the raw (non-normalised) dataset is 164 Hz. If we ask how this compares to the threshold for *perceptual* significance, one index of a possible answer is the range of typical values guaranteeing reliable perceptibility in experimentation. Cross-linguistic work on formant-frequency discrimination (Kewley-Port and Watson, 1994; Kewley-Port et al., 2005) suggests that the baseline  $\Delta F1$  required for reliable perceptual discrimination is in the region of 20Hz, so effects greater than around  $z \pm 0.12$  over our whole dataset are plausibly perceptually meaningful and interesting.

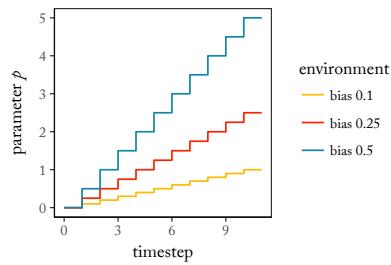
#### 4.2 Triggering segments

**FIXME Keep all this information, but reorganise — data first, then make the point that it's rhotic triggered, rhotic loses 'control' relatively quickly, and rate of change / degree of change is not proportional to phonetic precursor-ness**

In section 2.3.2, we considered the extent to which the individual sonorants constitute robust phonetic precursors to vowel lowering. The answer is not entirely conclusive; but the strength of the precursor in each environment does not seem to be constant, and the cross-linguistic evidence is variable. Further phonetic discussion of the case of the Turkish lateral appears in section 4.2.2 below; we argue therein that the lateral in Turkish does not represent a strong acoustic-articulatory trigger for vowel lowering.

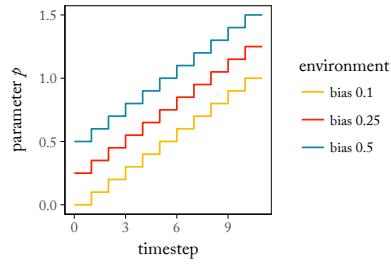
If it is really the case that different coda sonorants constitute different sizes of trigger, then we might well expect a difference in within-context realisations of pre-sonorant /e ø/, further conditioned by the individual coda segment. If this difference is visible in the data, how should it be understood? To recapitulate a little of the discussion in ?? with a visual example, models of sound change as hypocorrecting change

(Ohala, 1981) should predict that precursors of different strengths *induce effects of different sizes* and thus cause change to take place at different rates across the different coda contexts; as Fruehwald (2016), hypocorrective changes proceed proportionally to the sizes of their precursors. Assume a hypothetical parameter  $p$ , which varies over time in several different contexts; at each time step,  $p$  is incremented by an amount corresponding to a contextual bias favouring change. For contexts corresponding to biases of 0.1, 0.25, and 0.5 respectively, the parameter's evolution over time appears in figure 11 (after Fruehwald 2016, 378); after a few steps, values across contexts are further apart than they were initially.



**Fig. 11** A prototype of a *purely hypocorrecting* change: different biases cause unequal incrementation and different end realisations of phonetically-controlled parameter  $p$ . Conditions:  $p$  starts at 0, and is incremented by the value of the bias at each time step.

There are then multiple possible interpretations if we find a real statistical effect differentiating one pre-sonorant context from another. If the pathway of change was the one schematised in figure 11, then a visible difference across coda sonorants represents the end-product (or the intermediate point) of a hypocorrecting change: if the mechanism of phonologisation *remains* strictly hypocorrecting throughout, then we expect the effect of the initial difference in bias strength to be linearly multiplied over time. The alternate hypothesis is that the incrementation of phonological change is not strictly proportional to the strength of contextual phonetic biases; any visible difference across coda sonorants should be taken to represent *the original* triggering bias, but not the incrementation and compounding of that bias. If the fictional parameter in figure 11 is subject to an *initial* ‘phonetic’ bias differentiating its values across contexts, but then undergoes a fixed increment at each time step, we arrive at the situation in figure 12.



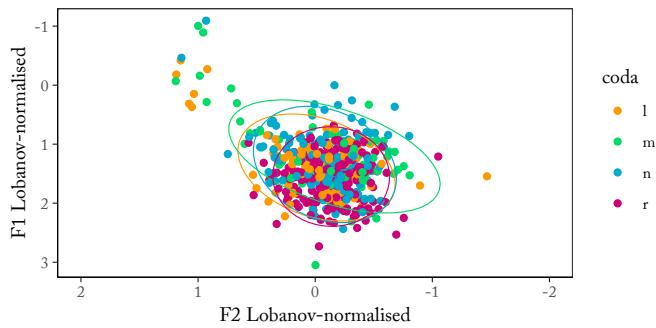
**Fig. 12** Change with a non-additive bias; we still see different end realisations of the phonetically-controlled parameter  $p$ , but the size of the contextual effect now remains constant. Initial conditions:  $p$  is 0.1, 0.25, 0.5 in the appropriate contexts, and is incremented by 0.1 at each time-step.

Within the set of sonorant triggers—the nasals, the lateral, and the rhotic—do further effects exist, either globally or subject to interspeaker variation? One possibility is that the final state of pre-sonorant /e/ and /ø/ realisations is further differentiated by coda segment; if this differentiation exists and corresponds to the ‘strength’ of the phonetic precursor provided by a given coda segment, this would fall in line with predictions about the end state of a hypocorrective, phonetically-motivated and bias-driven change (cf. ??), but the absence of either of these properties might point to a different pathway. In the case under consideration, one way to test which of these interpretations is more meaningful is to consider those tokens for which lowering is blocked by the placement of the syllable boundary. Recall (section 2.2) that resyllabification under affixation blocks lowering, as in [gi.zæm] ‘mystery’ but /gizem-i/ [gi.ze.mi] ‘mystery-dat’; in open syllables preceding an onset sonorant, we do not predict the existence of any systematic, grammatically-controlled lowering. If it is the case that there exists an effect differentiating tokens of /e/ in open syllables followed by *onset /r/* from other tokens of /e/ in non-final open syllables, we should find it reasonable to conclude that there is a phonetic effect of post-vocalic rhotics. If it is also the case that this effect is ‘similar’ in size to any effect distinguishing tokens preceding *coda* rhotics from those preceding other coda sonorants, then we suggest that the phonologisation of /e/-lowering has not in itself left behind the phonetic trace that pure-hypocorrecting accounts predict.

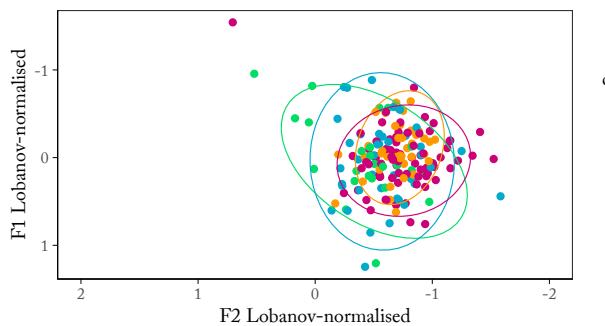
Consider the  $F2 \times F1$  scatterplots given in figures 13 and 14; these represent a subset of the data corresponding to all tokens followed by coda /r/, /l/, /m/, /n/ only. Each point represents a single pre-sonorant token of /e/ and /ø/ respectively, with 95% confidence ellipses indicating the spread of realisations. Inspection suggests that there is no substantial discontinuity and certainly no further categorical allophony; there is a slight predominance of pre-rhotic /e/ tokens at higher F1 values (lower, more open realisations).

Examining the  $F2 \times F1$  landscape after averaging and weighting by speaker (figure 15), however, suggests that the majority of this pre-rhotic effect can be attributed to productions from two speakers, F01 (Istanbul, b. 1997) and F04 (Izmir, b. 1988). F01 shows a generally higher F1 across contexts than other speakers, but all cross-context variation is well within confidence intervals and cannot be attributed to a ‘real’ effect. The separation between pre-rhotic tokens and other pre-sonorant tokens

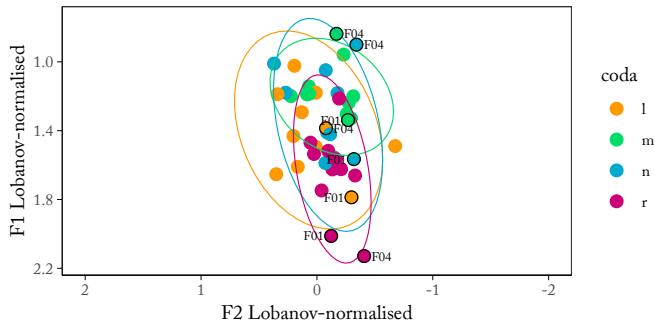
for F04 appears genuinely significant, and may represent an individual idiosyncracy in production. In the following section 4.2.1, statistical models are presented that account for interspeaker variation.



**Fig. 13**  $F_2 \times F_1$  space for pre-sonorant /e/, by individual coda segment. Each point corresponds to one vowel produced by a single speaker.



**Fig. 14**  $F_2 \times F_1$  space for pre-sonorant /ø/, by individual coda segment. Each point corresponds to one vowel produced by a single speaker.

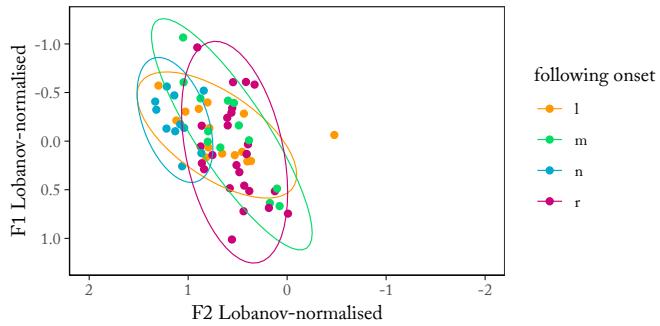


**Fig. 15**  $F_2 \times F_1$  space for pre-sonorant /e/, by individual coda segment. Each point corresponds to the *mean* of all corresponding tokens for a single speaker. Speakers F01 and F04, both of whom show particularly high F1 values in pre-rhotic contexts, are highlighted.

Consider further the scatterplot given in figure 16, representing all tokens of /e/ in *unstressed open syllables* immediately followed by an onset sonorant across the syllable boundary. Although the number of tokens is not large, there is some evidence to suggest that there is a meaningful effect of the rhotic on a vowel that precedes it *across* a syllable boundary. Vowels in open syllables are very susceptible to other height effects (sections 4.3 and 4.3.1); with this said, even though the conditions for raising effects are met in many of the pre-sonorant tokens here, their distribution is even across the set of lexical items that appears in figure 16, which appears below in table 9. Contexts in which the *raising* of section 4.3.1 might be expected to apply are indicated by shaded cells.

r	l	n	m
be.ra.ber	ka.le.ler	fe.ni	er.de.mi
def.te.ri	ge.lir	men.ge.ne	el.le.mek
sek.re.ter	ker.ten.ke.le	ge.nel	te.mel

**Table 9** Items in the test set containing open syllable /e/ followed by a sonorant onset.



**Fig. 16**  $F_2 \times F_1$  space for /e/ in *open syllables*, by the content of the immediately following syllable onset. Each point corresponds to one vowel produced by a single speaker.

#### 4.2.1 Modeling

Do individual [sonorant] segments in fact trigger different degrees of /e/-lowering or /ø/-lowering, despite the unclear state of such effects on visual inspection? A more reliable route to an evaluation of the significance of particular coda environments is modeling: as before, and for similar reasons, we fit two further mixed-effects models—for /e/ and /ø/ respectively—to a subset of the data consisting solely of those tokens with a coda sonorant in the right-hand environment, with F1 as dependent variable, coda (m, n, l, r) and stress ('stressed' or 'unstressed') as categorical fixed effects. Summaries of these results are presented in tables 10 and 11, including confidence intervals calculated via semi-parametric bootstrap estimation; the base levels assumed by the model are 'coda /l/' and 'unstressed'.

The model in table 10 does not support a claim that a significant difference exists between *pre-lateral* /e/ and *pre-rhotic* /e/. Realisations preceding *nasals* are generally higher (lower values of F1); what cannot be diagnosed from this analysis is whether this represents a systematic property of the phonologically-conditioned environments, or a 'top-level' effect of nasalisation in pre-nasal vowels (driven by the presence of nasal anti-formants).

Parameter	Estimate	95% confidence	Bootstrap
Intercept: coda /l/	1.47751	[1.230, 1.726]	
Coda: /m/*	-0.44463	[-0.7648, -0.1452]	
Coda: /n/*	-0.33914	[-0.6501, -0.0239]	
Coda: /r/	0.08575	[-0.1924, 0.3616]	
Stress: stressed*	-0.14399	[-0.4413, 0.1428]	
coda /m/ × stressed	0.21697	[-0.2289, 0.6633]	
coda /n/ × stressed	0.37876	[-0.0244, 0.7911]	
coda /r/ × stressed	0.08133	[-0.2982, 0.4731]	

**Table 10** Estimated means and derived confidence intervals for fixed effects in the model  $F1 \sim \text{coda} \times \text{stress} + (1 + \text{coda} + \text{stress}|\text{speaker}) + (1|\text{word})$  for /e/. 95% confidence intervals are calculated on 6,000 semi-parametric bootstrap estimates (R: `bootMer` from `lme4`), for which density plots are shown in the final column.

Parameter	Estimate	95% confidence	Bootstrap
Intercept: coda /l/	-0.21494	[-0.4132, -0.0175]	
Coda: /m/	0.22929	[-0.0273, 0.4198]	
Coda: /n/	0.13411	[-0.0567, 0.3241]	
Coda: /r/*	0.20231	[0.0249, 0.3791]	
Stress: stressed	0.20982	[-0.0164, 0.4321]	
coda /m/ × stressed	<i>no tokens, dropped from model</i>		
coda /n/ × stressed	0.14977	[-0.1268, 0.4321]	
coda /r/ × stressed	-0.13339	[-0.3727, 0.1141]	

**Table 11** Estimated means and derived confidence intervals for fixed effects in the model  $F1 \sim \text{coda} \times \text{stress} + (1 + \text{coda} + \text{stress}|\text{speaker}) + (1|\text{word})$  for /ə/. 95% confidence intervals are calculated on 6,000 semi-parametric bootstrap estimates (R: `bootMer` from `lme4`), for which density plots are shown in the final column.

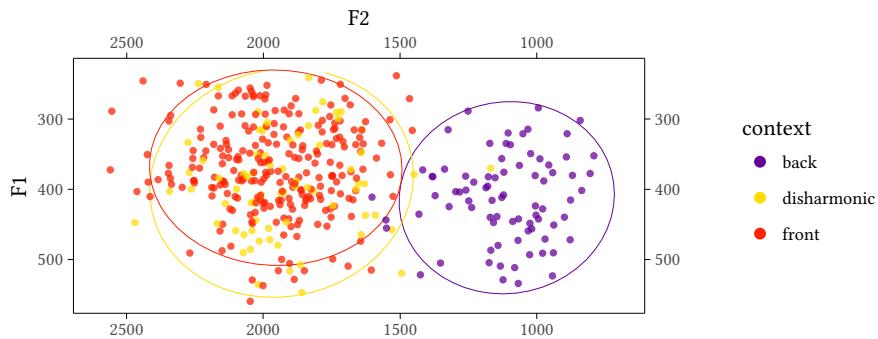
#### 4.2.2 The lateral

In section 2.3, we discussed possible motivations, phonetic and structural, for anticipatory F1-lowering, with the note that the status of the lateral would be addressed in greater depth in the following sections. What we are interested in here is whether or not the lateral itself constitutes an independently-motivated trigger for a vowel height effect: recall the discussion in ????, in which it was proposed that the analysis of the ‘naturalness’ of active classes in phonological change should take into account the

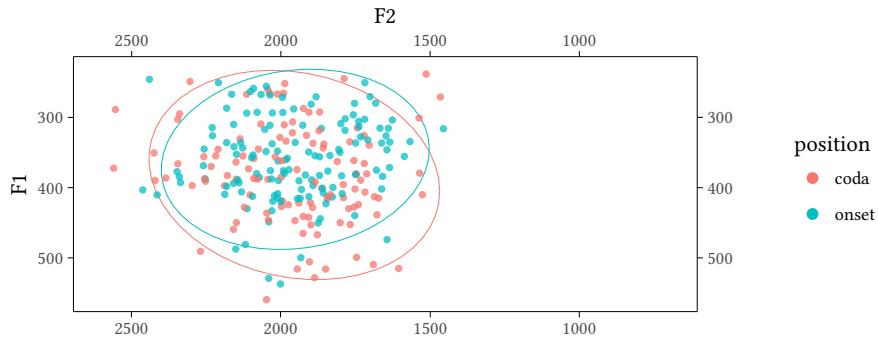
individual likelihood that each segment in that class constitutes a plausible phonetic precursor to the change in question, rather than the overall structure of the class alone.

I reference in section 2.1 the descriptive claim that the Turkish /l/ has an allophonic distribution conditioned by the backness of the surrounding vowel. The expected status of the lateral as a phonetic precursor to height effects then has a necessary dependence on its status as plain, velarised, or palatalised; these impressionistic categories correspond to a significant degree of acoustic variation (Proctor 2009 for an overview). Acoustic studies of various laterals suggest a correspondence between the first two formants and the degree of ‘darkness’, with a larger  $F2 \times F1$  difference corresponding to a ‘lighter’ (and thus further fronted) [l] (Sproat and Fujimura, 1993; Turton, 2014; Carter, 2002; Recasens and Espinosa, 2005). The locus of this difference is largely in  $F2$  variation, with  $F1$  relatively invariant throughout even in dynamic analysis (Strycharczuk and Scobbie, 2016), but due to the non-independence of  $F1$  and  $F2$  a correspondence is expected between backing (decrease) in  $F2$  and openness (increase) in  $F1$ ; Sproat and Fujimura (1993); Carter and Local (2007) give increased  $F1$  as a secondary correlate of /l/-darkness. If anticipatory coarticulation with a lateral affects the quality of the vowel immediately prior, we may then hypothesise that the major effect of a *velarised* lateral on a front vowel, corresponding in gestural terms to predorsum lowering and postdorsum retraction (Recasens, 2012), would be reflected in a  $F2$  decrease from the V target towards the lower /l/-target. It is less clear that a fronted lateral should cause an increase in  $F1$  as an anticipatory coarticulatory effect, since the predicted transition into such a lateral from a mid vowel involves a drop in  $F1$  and a sharp increase in  $F2$ ; indeed, the tongue-tip gesture corresponding to alveolarity corresponds rather to predicted raising, and thus  $F1$ -decrease. Recasens (2014) gives /e/-lowering to [a] post-palatal lateral in Catalan as a possible progressive dissimilatory process in a very small handful of cases, but gives pre-lateral raising as far more frequent. **FIXME Diskin et al lowering in Melbourne — see if anything relevant to cite**

The remarks in this section constitute a very preliminary study of the acoustic properties of the Turkish lateral; the intent is not at all to provide a full study, but to characterise in broad strokes the expected nature of the transition in /el/: can /e/-lowering in /el/ admit a simple coarticulatory explanation? In order to test this, all instances of /l/ were extracted from the production data collected in this study, for measurement. Across 11 (female) speakers, this yielded 627 tokens for analysis; of these, 53 were discarded due to regressive manner assimilations (to coronals s, z, n) or due to phrase-final devoicing and frication leading to unreliable formant measurements. For the remaining 574 datapoints,  $F1$  and  $F2$  were measured at the midpoint, to give a rough indicator of the ‘frontness’ of the lateral. A summary visualisation appears in figure 17; tokens were coded ‘front’, ‘back’, and ‘disharmonic’ by adjacent vowel (respectively: exclusively i, y, e, ø; exclusively u, ɯ, ʊ, a, o; one vowel from both sets).



**Fig. 17**  $F_2 \times F_1$  space for the lateral /l/, measured at midpoint, by vocalic environment: ‘disharmonic’ tokens are adjacent in the string to both a front vowel (i, e ~ æ, y, ø ~ œ) and a back vowel (ɯ, u, o, a), within a single syllable or across the syllable boundary.



**Fig. 18**  $F_2 \times F_1$  space for *fronted* tokens of lateral /l/, measured at midpoint, by position.

The distribution of the Turkish laterals in  $F_2 \times F_1$  space thus appears discontinuous and categorical, conditioned by the adjacent vowels and with no particular dependence on syllabic position. Without exception, the laterals adjacent to (either preceded by or preceding) front vowels have significantly raised  $F_2$ , including those occupying codas immediately following phonologically lowered /e/ [æ]. Other plausible predictors (vowel quality, stress) had no statistically meaningful effect.

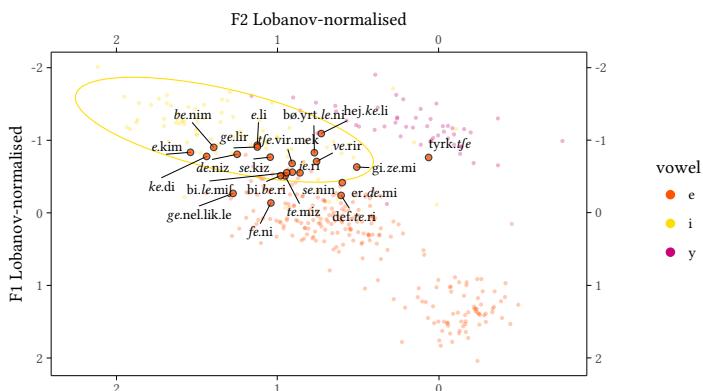
Why are the realisational details of the Turkish coda lateral interesting to us? In a language like English, in which the distribution of the plain lateral and the heavily velarised lateral is strictly conditioned by position in the syllable, the velarised coda lateral becomes a natural and well-motivated phonetic trigger for vocalic effects. In any of the contexts in which we are interested in the Turkish lateral, however, *no such trigger exists*.

### 4.3 Duration

One possible origin for the vowel-lowering alternations that we see is in coda-conditioned loss or gain of duration, as in the analyses of vowel quality effects that we cover in section 2.3.1. If this is the case, then we might expect some systematic pattern to appear in the sample; both in the duration-F1 correlation, and in the predicted duration for each phonologically-active set of environments (pre-obstruent; pre-sonorant; open).

#### 4.3.1 /i/-triggered height effects

Recall from tables 3 and 7 that there were statistically-significant estimates distinguishing /e/-realisations in *stressed* and *unstressed* open syllables, and in *stressed* syllables for each of the three environment types, but that we could not reject the zero-difference hypothesis for ‘obstruent’ and ‘open’ in unstressed syllables. In terms of our estimated parameters, this is as follows:  $F1_{\text{obstruent}} \sim F1_{\text{open}} < F1_{\text{sonorant}}$  in unstressed syllables, but  $F1_{\text{obstruent}} < F1_{\text{open}} < F1_{\text{sonorant}}$  in unstressed syllables. Both cases for /e/ are distinct from the case for unstressed /o/, for which  $F1_{\text{open}} < F1_{\text{obstruent}} < F1_{\text{sonorant}}$ . One possibility is that the state of unstressed /e/ is in fact a *superposition* of an /o/-like case (open < obstruent < sonorant) and a stressed-/e/-like case (obstruent < open < sonorant).



**Fig. 19** F2×F1 space for /e/, with the high vowels /i/ and /y/. Ellipse: 95% confidence region for tokens of /i/. One point corresponds to the *mean* across all tokens of a single word.

Consider the F2×F1 space depicted in figure 19 above. All items of the form /(C)e.Ci/ are highlighted and transcribed; /e/-realisations in these contexts fall within the bounds of tolerance for the high vowel /i/, and skew the overall F1-distribution for /e/ in open syllables. This then suggests that there is something special about the unstressed open syllables; we refrain from answering the resultant questions now, but the problem will recur in sections 4.3 and 5.3.

### 4.3.2 The duration-height connection

In section 2.3.1, we devoted some attention to the relationship between duration and vowel quality; what we did not address in full was the state of the debate as to whether this relationship has a purely mechanical etiology, or instead arises from the variable phonologisation of durational targets. The mechanical explanation, rooted solely in facts of articulation and low-level phonetics, makes several fundamentally testable predictions. If low vowels are longer due to the length of the jaw-opening gesture (Lehiste, 1970), then the ‘extra’ length must be particularly visible in the onset and offset formant movements towards the low vowel (Lisker, 1974); if instead the steady-state is the locus of increased duration, this argues for a phonological explanation. If durational effects are purely mechanical, then they should remain constant across speech rates; Solé and Ohala (2010) find that this is indeed the case in Japanese, but is not the case in Catalan or English, and argue therefore that the correlation between F1 and duration is phonologically-controlled in the latter languages but phonetically-controlled in the former. Two further predictions of the pure-phonetics account are especially relevant to the case in this paper, and appear in example 9.

*Example 9* Given a strictly phonetic account of the F1–duration correlation:

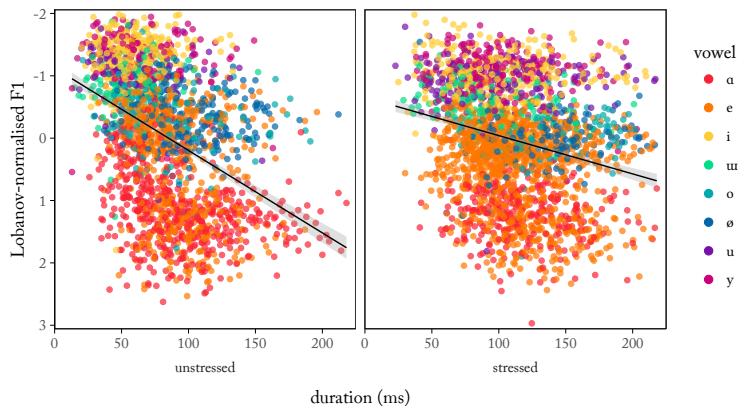
1. A positive correlation between F1 and duration should hold both across categories and within categories; that is, given multiple tokens of a single vowel, it should be the case that the highest instances thereof are shorter than the lowest instances.
  - This effect does not have to be constant in magnitude across categories. The lowest and highest vowels are inherently constrained, and therefore there should be durational ‘ceilings’ at the extremes of these categories; duration itself cannot approach 0 very closely, or approach arbitrarily large values. It’s therefore likely that there is non-linearity at both the extremes of the vowel space and the extremes of duration, even if the overall relationship is linear.
2. If realisations of an individual vowel systematically shift in height as the result of phonological change, we should expect an attendant pattern in duration.

There is reasonable evidence that these particular predictions do not hold in all cases. Tauberer and Evanini (2009) show, for various sound changes in North American English, that the distributions of dialectal variation in F1 and in duration appear independent; that is, duration is not affected by phonological changes manipulating vowel height. Toivonen et al. (2015) find, for English and Swedish, that although there is a cross-category correlation between height and duration (tokens of phonemically high vowels /i/ are longer than tokens of phonemically low vowels /æ a/) this does not always persist within categories: higher and lower realisations of [ɪ] show no systematic patterning in duration. The alternate explanation for any apparent positive correlation between duration and F1 is then a phonologically-controlled one: each vowel (category) has an independent phonologised duration target, and it happens to be the case that these targets are shorter for high vowels than they are for low vowels. In a sense, though, a solution of this type just moves the problem to a slightly different tier of analysis: why should this separation in durational targets arise and phonologise? The answer that Solé and Ohala (2010) propose essentially implies that

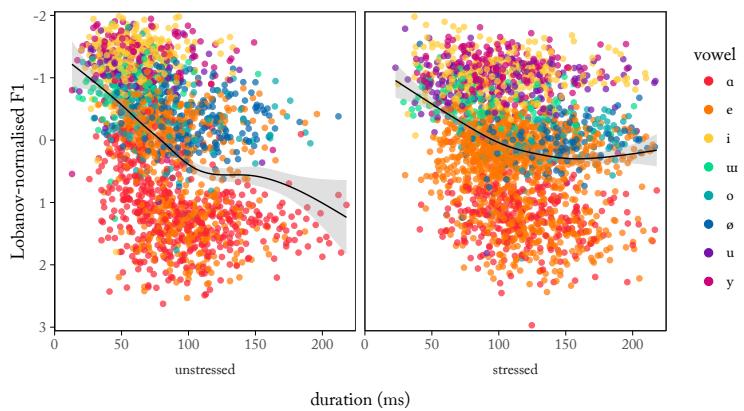
phonologisation ‘overrides’ what is presumed to be the original mechanical bias: the use of duration as a marker of phonological identity is ultimately phonologised from an uncontrolled phonetic preference for long low vowels, and short high ones.

These points then give rise to several nested questions about the experimental data in this paper. The broadest questions are those pertaining simply to the account itself: is there a general positive correlation, across all vowels, between duration and the first formant? Does this hold for within-category behaviour? The subtler issue is that of the relationship between duration and the phonologised height pattern that we have established throughout this section: given that the end-point of sound change in Turkish is a systematic difference in mid-vowel height constrained by environment, do we see an attendant distribution of duration across these environments? If not, how do we interpret the mismatch? We are also really asking here about the direction of the causal relationship between the first formant and duration; do changes in F1 drive responses in duration, or vice versa—or do they feed into each other in some respect?

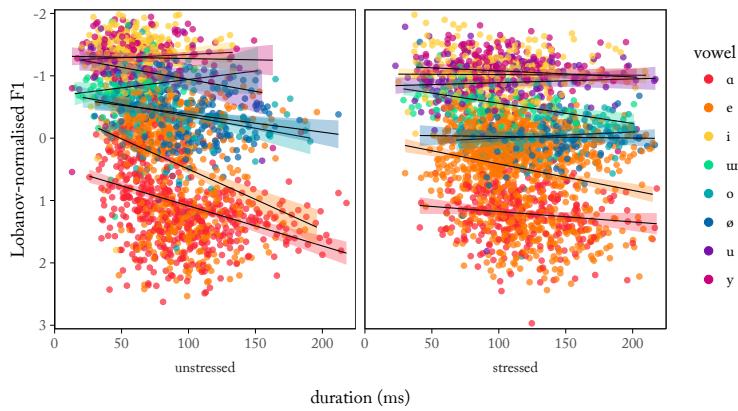
In this subsection, we take a filtered dataset corresponding to all ‘reasonable’ measurements of duration: points more than 3 standard deviations (39.75ms) away from the dataset mean of 99.04ms were dropped, giving an adjusted range of [12, 219] ms. (This corresponded to a loss of a little under 5% of the data, 214 points of 4747 total; this can essentially be attributed to automated measurement errors.) Consider the illustrations in figures 20–23. We have split visualisations by stress throughout to ensure that we are sensitive to the unquantified effect of stress on duration. What figure 20 then represents is the extent to which an across-the-board, cross-category correlation between F1 and duration exists, in a very naïve understanding that ignores confounds and intricacies. It does seem that there is a correlation, and that it is statistically non-insignificant; but at the same time, it explains relatively little of the variation in the dataset, and it is unclear what its origin is. Figure 21 fits non-linear functions to the data instead, and suggests that the greatest deviations from the pure-linear model do indeed arise at the edges of the space, and in particular are driven by the longest vowels.



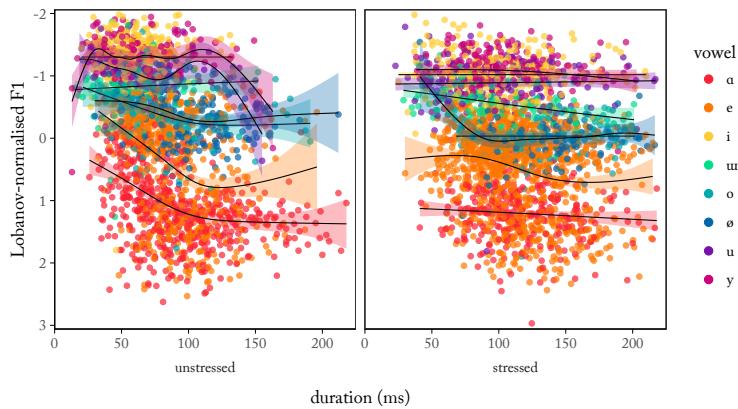
**Fig. 20** All points in the sample: Lobanov-normalised F1 plotted against duration in milliseconds, with each point representing one token and coloured by vowel category; panels are split by the presence or absence of stress on the corresponding syllable. The line in each panel represents the best linear-model fit for  $F1 \sim \text{duration}$ , with no further predictors. Tests of correlation: *unstressed* Pearson's  $r = 0.4057$ ,  $t 20.103$  on 2051 degrees of freedom  $p < 2.2e-16$ ; *stressed* Pearson's  $r = 0.2418$ ,  $t 12.237$  on 2410 degrees of freedom  $p < 2.2e-16$ .



**Fig. 21** All points in the sample: Lobanov-normalised F1 plotted against duration in milliseconds, with each point representing one token and coloured by vowel category; panels are split by the presence or absence of stress on the corresponding syllable. The line in each panel represents the best *cubic spline* fit for  $F1 \sim \text{duration}$ , with no further predictors.



**Fig. 22** All points in the sample: Lobanov-normalised F1 plotted against duration in milliseconds, with each point representing one token and coloured by vowel category; panels are split by the presence or absence of stress on the corresponding syllable. Smoothing lines represent the best linear-model fit for  $F1 \sim \text{duration}$ , for each vowel. Correlation statistics: table 12.



**Fig. 23** All points in the sample: Lobanov-normalised F1 plotted against duration in milliseconds, with each point representing one token and coloured by vowel category; panels are split by the presence or absence of stress on the corresponding syllable. Smoothing lines represent the best cubic spline for  $F1 \sim \text{duration}$ , for each vowel.

In theory, if a purely phonetic and non-cognitively-controlled effect applies, it should persist at all levels and, as example 9, apply both within phonemic categories and across them. Consider the linear relationships within individual vowel categories (but *not* further split by following environment) shown graphically in figure 22, and listed in table 12; these represent tests of  $F1 \sim \text{duration}$  over tokens corresponding to each vowel category and stress placement, under the assumption that the relationship is linear and has no other major dependencies. Results are presented with the

Pearson correlation coefficient  $r^{21}$  as a measure of the relationship between variables, and with  $t$ -distribution tests for significance; significant correlations are highlighted. In the table below, the  $t$ -test's degrees of freedom (column  $df$ ) indicate, as always, the number of tokens measured in each category; what these results then suggest is that the apparent cross-category correlation between F1 and duration that we might hypothesise from figure 20 may well be driven by the numerical over-representation of a few vowel categories.

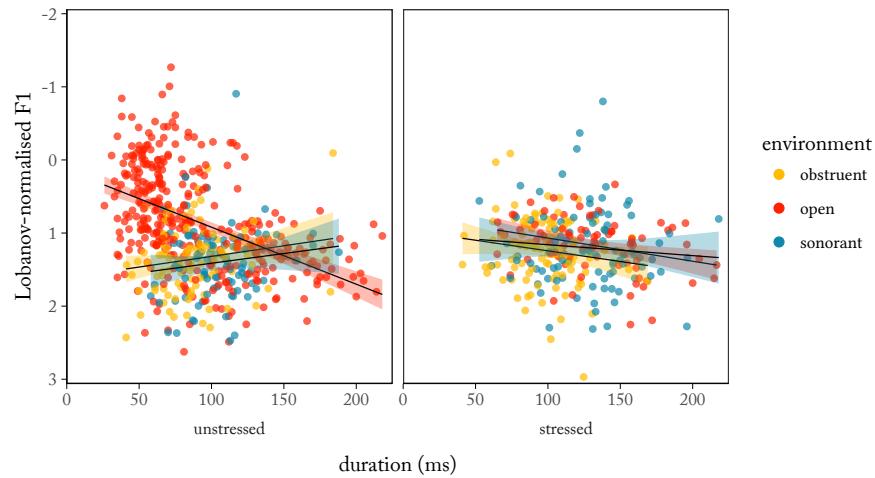
Vowel	stress	Pearson's $r$	$t$	df	$p$
a	unstressed	0.357	8.91	543	7.69e-18
	stressed	0.120	2.02	283	0.044
e	unstressed	0.286	6.96	544	1.01e-11
	stressed	0.211	7.02	1057	4.04e-12
i	unstressed	-0.063	-1.03	262	0.304
	stressed	0.010	0.172	295	0.863
u	unstressed	-0.138	-1.48	112	0.140
	stressed	0.274	3.427	145	0.001
o	unstressed	0.281	3.332	130	0.001
	stressed	-0.087	-0.837	91	0.404
ø	unstressed	0.194	2.994	230	0.003
	stressed	0.025	0.207	145	0.759
u	unstressed	0.238	2.176	79	0.033
	stressed	-0.062	-0.797	164	0.427
y	unstressed	0.028	0.331	137	0.741
	stressed	0.063	0.931	221	0.353

**Table 12** Tests of correlation for  $F1 \sim \text{duration}$ , within individual vowel categories. Cases for which a 5% criterion of significance is met are highlighted.

If the apparent dependence of F1 on duration is really dominated by uneven numbers of tokens across vowel categories, then what it is useful to have is a clearer understanding of the internal structure of those categories. Does the relationship between duration and height within a single category submit to further sub-segmentation? Is it possible to establish a relationship between duration itself and the other phonological conditioning that applies within a single vocalic category?

Consider the indicative figure given in figure 24, which represents the best-fit linear model for  $F1 \sim \text{duration}$  for tokens of underlying /a/, subdivided further by environment: this suggests that the apparent F1-duration relationship for /a/ lies almost entirely in unstressed (non-final) open syllables.

<sup>21</sup> This ranges in ideal cases between  $[-1, 1]$ .



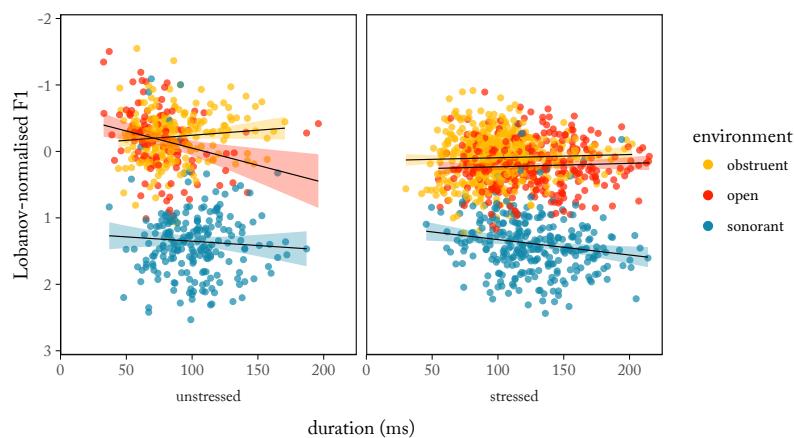
**Fig. 24**  $F1 \sim \text{duration}$ : for all tokens of /a/ in the sample, Lobanov-normalised F1 plotted against duration in milliseconds, with each line representing the best linear-model fit for tokens in the corresponding syllabic environment.

To reiterate slightly before considering the cases of /e/ and /ø/, there are really two possible scenarios in which duration is significant to the phonologisation of mid vowel lowering. One is the case in which a  $F1 \sim \text{duration}$  correlation persists down to the level of the individual vowel category and below—that is, holds also for each analytic sub-category (environment of interest)—which implies a relationship implemented at a fairly low, likely articulatory level, although this is already somewhat countered by figure 22. Another is the case in which  $F1 \sim \text{duration}$  fails to hold in this way, but  $\text{duration} \sim \text{environment}$  does: that is, even though  $F1$  is not ‘really’ distributed by duration, duration itself (for the phonologically-targeted vowels /e ø/) has substantially different expected values across the different environments. If neither of these holds, or if we find some more complex relationship between duration, height, and context, then we arrive at a set of related questions; are durational effects under the control of the phonetics or the phonology, and is there a systematic way in which they might be related to effects on height?

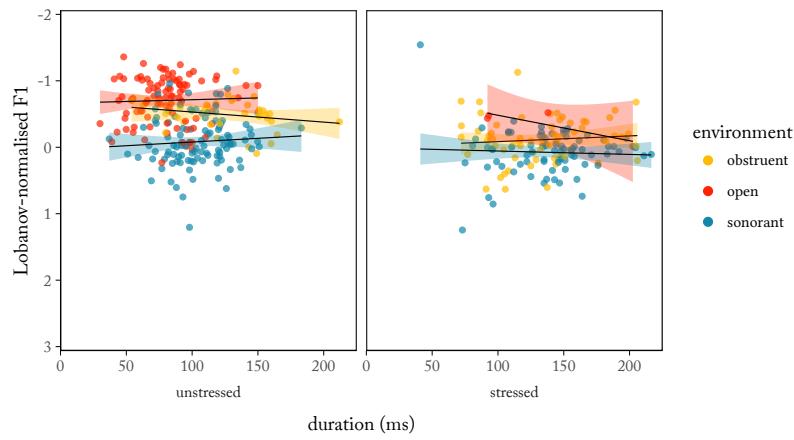
Consider the visualisations in figures 25 and 26. We can attempt to draw linear fits through the data, but their performance isn’t particularly good; as for /a/, the only statistically significant positive correlation between  $F1$  and duration for /e/ appears in unstressed open syllables. What this seems to imply is that there is indeed a relationship between duration and height, but that its ontology is more complex than the gestural account suggests.

*Unstressed vowel reduction* Recall section 4.3.1, in which tokens of /e/ in unstressed open syllables appeared to be raised if followed by /i/. The durational effect shown in figure 25 persists even if tokens of this type are excluded from the sample; that is, persists in all unstressed open syllables independent of following context. The reason

that this is interesting is then the long history of proposals connecting (or disconnecting) phonology and phonetics in vowel reduction (Barnes, 2007; Iosad, 2012); is there a systematic effect that we can characterise in either or both of /e Θ/, and can it be related back to the sonorant-conditioned pattern? The best answer that we can give to this question is a little speculative. There is no such effect for /Θ/; but recall from figure 24 the substantial relationship between height and duration for /a/ in unstressed open syllables, which was large enough to drive the false appearance of a cross-category effect across all vowels. The best interim conclusion is that in line with Tauberer and Evanini (2009); Solé and Ohala (2010); Toivonen et al. (2015), we should take seriously the idea that correlative relationships between durational targets and height targets are systematically set by the grammar; learners are perhaps biased towards the acquisition of a unidirectional relationship (that is, towards the general predictions of physiology), but it need not be the case that this relationship exists in the same state for every category, even within a single language.



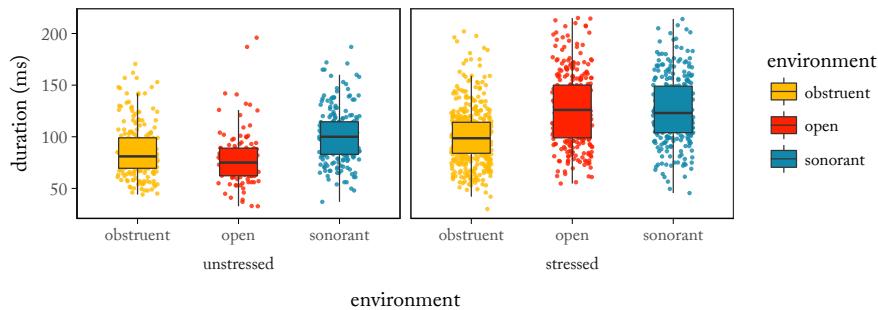
**Fig. 25**  $F1 \sim \text{duration}$ : for all tokens of /e/ in the sample, Lobanov-normalised F1 plotted against duration in milliseconds, with each line representing the best linear-model fit for tokens in the corresponding syllabic environment.



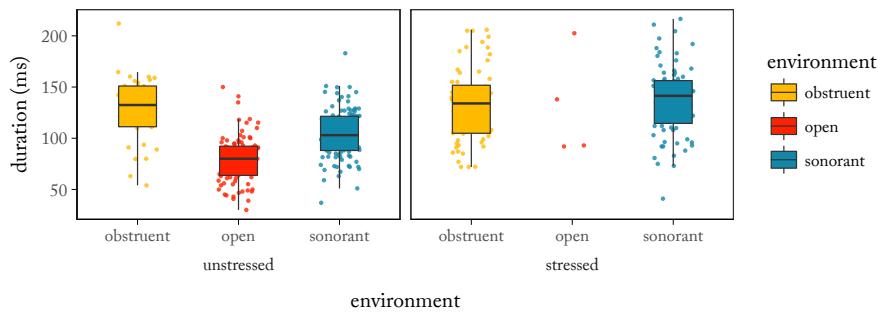
**Fig. 26**  $F1 \sim \text{duration}$ : for all tokens of /ø/ in the sample, Lobanov-normalised F1 plotted against duration in milliseconds, with each line representing the best linear-model fit for tokens in the corresponding syllabic environment.

#### 4.3.3 Dependencies within the mid vowels

Setting aside the issue of the relationship between duration and *height*, is there a systematic effect of conditioning environment on the duration of an individual vowel? Even in the absence of within-category correlation between duration and height, significant cross-category differences in duration can still exist, and can still represent some form of systematic relationship between conditioning environment and target. Cross-category significance would also rescue an account of the phonologisation of pre-sonorant lowering in which there is either some significant biasing pressure of duration, or some significant movement of durational targets in tandem with height targets. We plot per-category duration in figures 27 and 28, list the results of a linear mixed-effects model for per-category duration of /e/ in tables 13 and 14, and consider the potential for individual variation in figures 29 and 31.



**Fig. 27** Box-plot of duration, by conditioning coda environment and split by stress, for all tokens of /e/ in the sample.



**Fig. 28** Box-plot of duration, by conditioning coda environment and split by stress, for all tokens of /ə/ in the sample.

As before, a linear mixed-effects model was chosen in order to mitigate some of the effect of inter-speaker variation and idiosyncrasy on the result. The impressionistic suggestion of figures 27 and 28 is that although the state of the within-category linearity between F1 and duration is inconsistent, it is reasonable to posit that the overall distribution of duration is categorically-predicted: it is the case that overall, vowels are shorter in unstressed open syllables, but it is not necessarily the case that a predictive relationship holds between the within-category variation in vowel length and the within-category in height. The effect of category on the duration of /ə/ is particularly striking; the effect on the duration of /e/ is not as large in absolute units of time, but statistical modelling suggests that it is significant, and that there is an inversion of the expected cross-category duration between /e/ and /ə/ example 10.

Parameter	Estimate	95% confidence	Bootstrap
Intercept: obstruent*	93.452	[86.22, 100.52]	
Environment: open*	-10.945	[-20.94, -0.25]	
Environment: sonorant*	13.134	[3.78, 22.48]	
Stress: stressed	6.582	[-1.45, 14.85]	
open × stressed*	44.144	[36.00, 52.02]	
sonorant × stressed*	14.887	[6.45, 23.69]	

**Table 13** Estimated means and derived confidence intervals for fixed effects in the model  $\text{duration} \sim \text{environment} \times \text{stress} + (1 + \text{environment} + \text{stress} | \text{speaker}) + (1 | \text{word})$  for /e/. 95% confidence intervals are calculated on 6,000 semi-parametric bootstrap estimates (R: `bootMer` from `lme4`), for which density plots are shown in the final column.

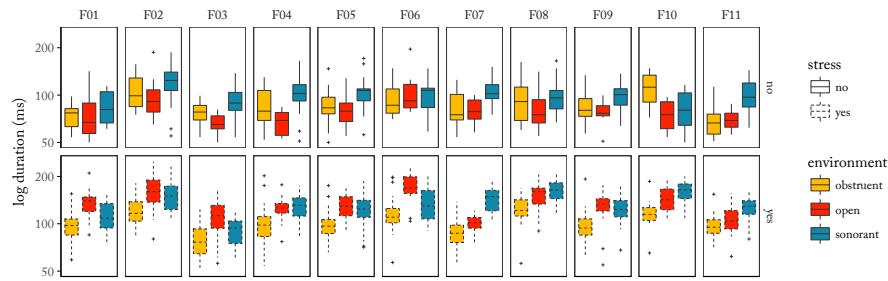
Group		SD	Correlation
<i>Word</i>	Intercept: obstruent	16.000	
	Environment: open	16.208	-0.85
	Environment: sonorant	20.439	-0.72 0.94
<i>Speaker</i>	Intercept: obstruent	7.521	
	Environment: open	12.406	0.31
	Environment: sonorant	9.262	0.07 -0.60
	Stress: yes	10.906	0.51 0.55 0.03
<i>Residual</i>		20.694	

**Table 14** Random effect standard deviations and correlations, model  $\text{duration} \sim \text{environment} \times \text{stress} + (1 + \text{environment} + \text{stress}|\text{speaker}) + (1|\text{word})$  (table 13), /e/.

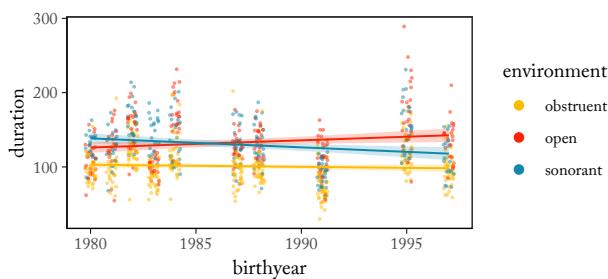
*Example 10* Conditioning environments, stress, and duration.

vowel	stress	duration by environment
/e/	unstressed	sonorant > obstruent > open
/ø/	unstressed	obstruent > sonorant > open
/e/	stressed	open > sonorant > obstruent
/ø/	stressed	sonorant ~ obstruent

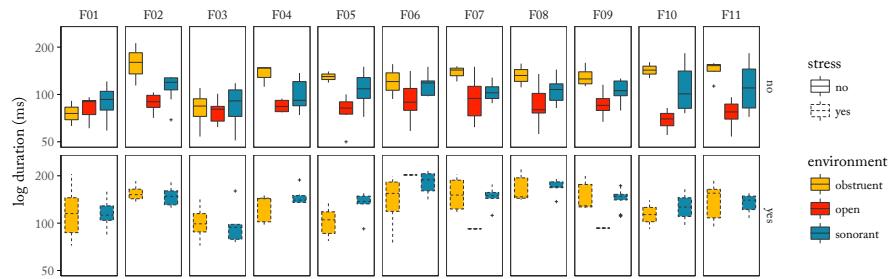
We can then ask whether this relationship continues to hold across individual speakers. The distributions of duration for each of the 11 female speakers in the sample are visualised in figures 29 and 31 for /e/ and /ø/ respectively. Qualitatively, there is no strong differential in the pattern visible across speakers for unstressed /e/, which is broadly concordant with the overall statement we give in example 10. The pattern in stressed /e/ is more variable, in the sense that there is an apparent reversal of the broad hierarchy of environments; for *older* speakers, like F08, F09, F10, and F11, the ordering of predicted durations of stressed /e/ runs sonorant > open > obstruent, which is counter to the behaviour of *younger* speakers for whom the prediction is instead open > sonorant > obstruent. Although the sample given in this paper suffers from a rather severe lack of time depth, as an indicative illustration we plot the relationship between year of birth and /e/-duration in figure 30. The pattern in /ø/ is more generally constant; there is no evidence for conditioning by age that meets statistical qualifications for significance, nor a particularly strong qualitative pattern, unless the youngest speaker's reversal of the typical ordering in unstressed syllables (F01) is to be taken as representative of a genuinely distinct system.



**Fig. 29** Duration in milliseconds, by individual speaker and split by stress, for /e/.



**Fig. 30** Duration in milliseconds, by speaker's year of birth, for stressed /e/.



**Fig. 31** Duration in milliseconds, by individual speaker and split by stress, for /ø/.

Does the variable but constrained patterning in duration reflect some artefact of the end-product of sound change in Turkish, or is it coincidental? One possible argument here is that the existence of patterning in duration is a reflection of the original conditions of phonologisation; for speakers who are behind in the change, a few traces of the original environment for phonologisation are seen, and these traces disappear for speakers who are further ahead. Due to the lack of age variation in the sample, this

is rather difficult to comment upon further; but the extent to which various apparent effects persist in a larger, and *older* sample, is a productive direction for an expansion of this work.

#### 4.4 Exceptionality

I have spent the previous sections arguing for generality and categoricity in the lowering of the Turkish mid vowels; the pattern is discontinuous in phonetic space, persists across a large test set and across all experimental participants, and varies under resyllabification in a manner consistent with phonologised positional restrictions. In this subsection, we remark briefly on the state of *exceptions* to these neat generalisations, which we divide into two broad categories; the first set consists of pre-sonorant non-undergoers of /e/-lowering, and the second, perhaps more interesting set of pre-obstruent *undergoers*: the latter is given in our sample largely by tokens of the aorist negative /-mAz/<sup>22</sup>, although it appears to apply more generally. These are presented in figures 32–36 below, with brief commentary.

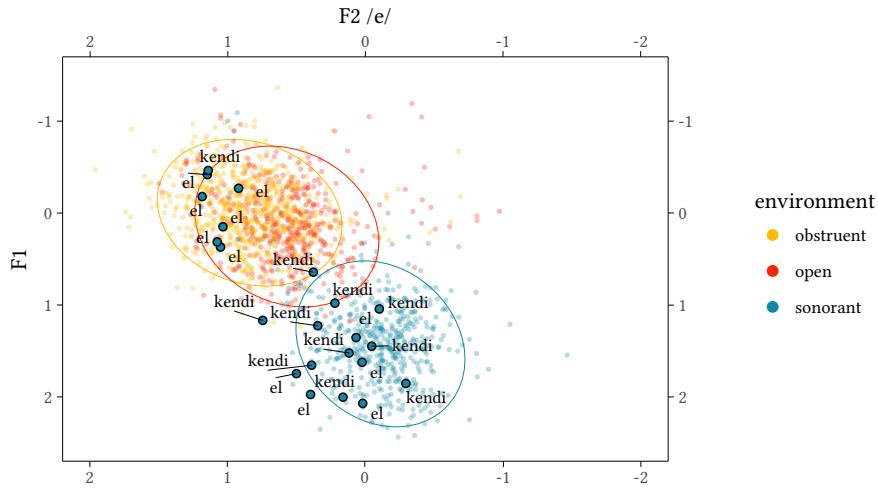
##### 4.4.1 Pre-sonorant exceptionality

Figure 32 shows, for *el* [el ~ æl] ‘hand’, and *kendi* [ken.di ~ kæn.di] ‘self’, a general instance of variability in relatively high-frequency items. Several further examples of exceptionality that seem to be more explicitly grammatical follow in figures 33–35; we present these only minimally here, as our experimentation did not test for exceptionality in sufficient detail to formulate robust hypotheses as to *what* that conditioning might be. The first, in figure 33, represents an arguably ‘false’ case of exceptionality; what appears initially to be unexplained opacity seems instead to arise from apparent ‘true’ geminates, syllabified as onsets rather than as coda-onset clusters; the lexical items involved have in common a status as comparatively-unadapted loans. This can be contrasted with /tel-li/ [tæl.li] ‘wired’, also shown, which presents a morphologically-doubled sonorant on the surface, but in which /e/-lowering does apply. Pre-sonorant lowering appears to be blocked for N+C clusters but not for R+C clusters, as figure 34. A final, particularly intractable exception appears in a poorly-characterised set of word-initial (and thus unstressed) sonorant-coda syllables, as figure 35 (data in this figure have been averaged by word and context; this is simply to keep visual clutter at manageable levels). There is no *general* constraint against lowering in a word-initial syllable: /erdem/ [ær.dæm] ‘virtue’, [kæn.di.mi.ze] ‘refl-1pl-dat’, and non-resyllabifying affixation does not induce exceptionality in an undergoer-root.

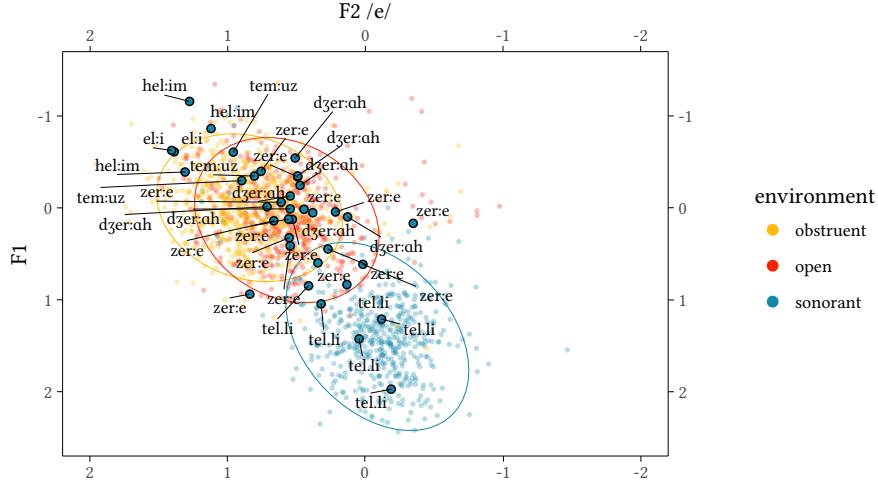
We are unable to find any evidence of exceptionality in /ø/, although due to the size of the test set and the greater incidence of variability in /ø/, this should be taken with a certain amount of salt. If it is genuinely the case that there is no real exceptionality for /ø/, then this is broadly consistent with a model in which the ‘phonological’ status of lowering in /e/ is further advanced, and thus amenable to effects violating strict

<sup>22</sup> /a/ here indicates an affix vowel whose content is unspecified, and thus in [back] harmony with the final vowel of the root may be either [a] or [e~æ] on the surface; as in ??, in which we applied the same convention throughout.

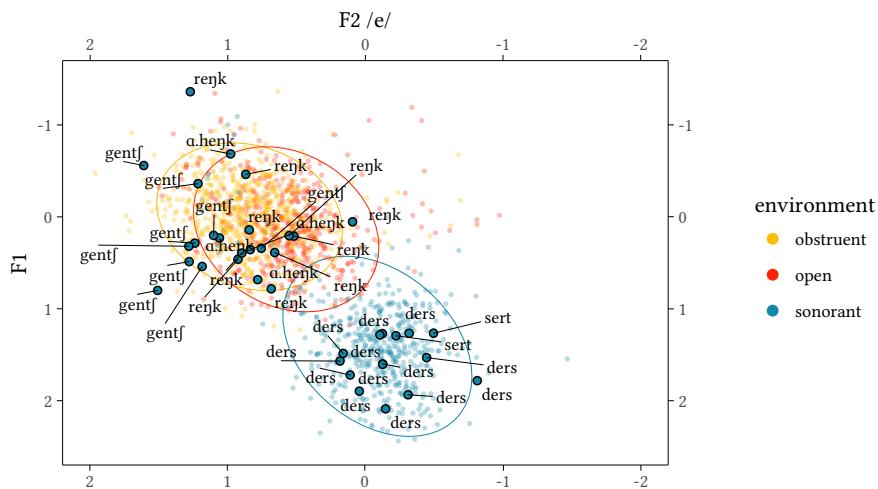
phonetic conditioning (see eg. Bermúdez-Otero 2015. If lowering generalises to /ø/ after /e/, we expect /ø/ to be behind in stabilisation).



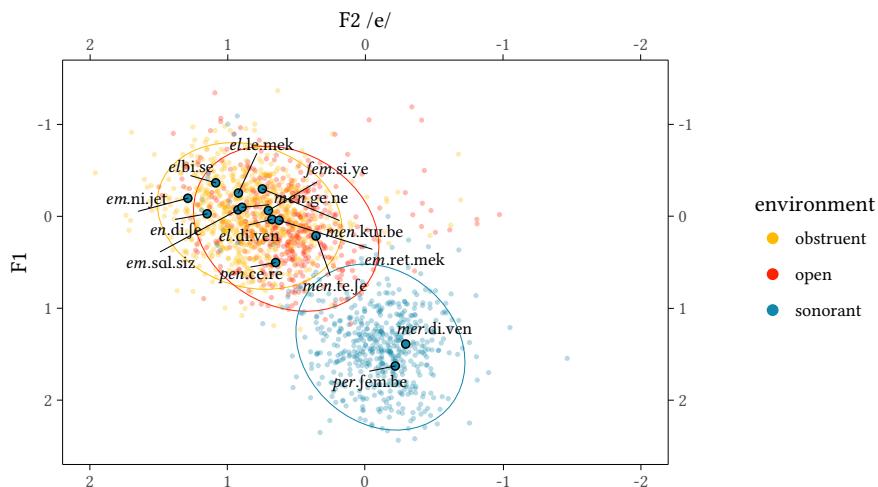
**Fig. 32**  $F_2 \times F_1$  space for all tokens of /e/, with apparently lexically-conditioned variability highlighted for *el* [el ~ æl] ‘hand’, and *kendi* [ken.di ~ kæn.di] ‘self’; each point represents the production of a single speaker.



**Fig. 33**  $F_2 \times F_1$  space for all tokens of /e/; exceptional ‘true geminates’ are highlighted and labelled, along with /tel-li/ [tel.li] ‘wired’ for comparison. One point represents one speaker’s production.

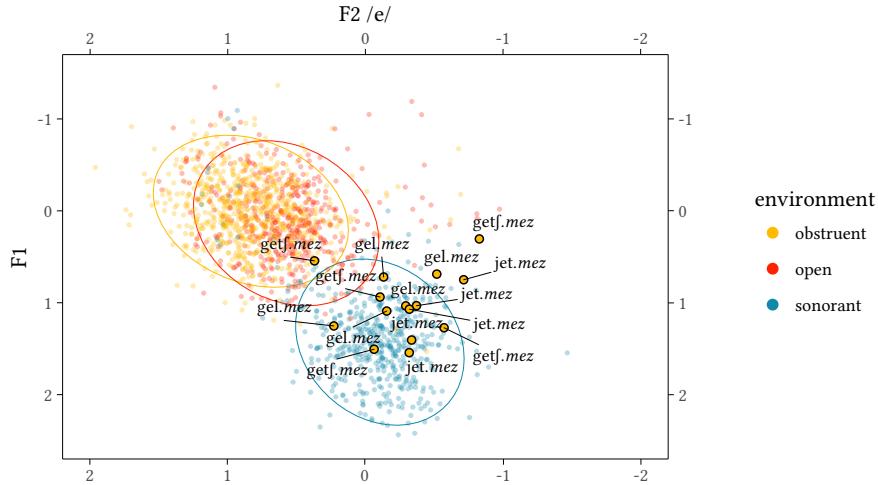


**Fig. 34**  $F_2 \times F_1$  space for all tokens of /e/; coda N+C clusters and coda R+C clusters are highlighted and labelled.



**Fig. 35**  $F_2 \times F_1$  space for /e/, all tokens shown in background. Highlighted points are averaged: one point in the space represents the average over all tokens for a target vowel in a particular word. Exceptionally-behaving initial syllables are marked, along with similar-seeming items whose patterning shows no corresponding exceptionality.

#### 4.5 Generalisations to the voiced fricatives



**Fig. 36** F2×F1 space for /e/, all tokens shown in background. Each highlighted point represents a single token, for a single speaker, of an item containing the negative aorist /mez/; either *gelmez* [gæl.mæz] ‘not come’, or *geçmez* [getʃ.mæz] ‘not go’.

Coda voiced fricatives in Turkish are relatively infrequent, for reasons possibly ultimately related to a significant rate of pre-pausal devoicing. Instances of the negative aorist /-mAz/ make up the majority of these, and the majority of /ez/ tested. In a set of 1,337,898 morphologically-complex types (parsed by Bilgin 2016, derived from the corpus of Sak et al. 2008), there were 91,798 <z>-final types, of which 2,104 were <ez>-final; of these, only 62 did not contain the negative aorist. (One of these others is *pekmez* [pek.mæz] ‘molasses’, which did appear in the test set and unproblematically showed /e/-lowering pre-z.)

The negative aorist /-mAz/ itself is both frequent and (perhaps optionally, although no speaker in our data showed a real exception) susceptible to the process of /e/-lowering—despite /z/’s non-membership in the natural class of [sonorant] segments to which the other triggers, irrespective of their independent status as strong precursors to F1 effects, belong.

What the existence of this piece of exceptionality then represents is a generalisation of what we have described as pre-sonorant lowering to at least one voiced obstruent. If this is the case, then we need to address some explanatory, descriptive, and diachronic problems—but quite deep ones, with the potential to reflect on many of the phonological issues that arise in this paper and in this thesis as a whole. Recall the cases presented in sections 5.1.1 and 5.1.2: in both Schaffhausen German and in Georgian, a single alternation applied to an active class that was a superset of some natural class. If mid-vowel lowering in Turkish applied exclusively before the

sonorants, then despite the apparent lack of phonetic cue in some pre-sonorant environments, we could diagnose it as having a set of environments whose members form a representationally-coherent natural class. With the inclusion of the voiced fricative /z/, this is no longer the case; an attempt to analyse the pathway of generalisation from actuating context to final set must therefore account for this movement across class boundaries. This informs the remarks that follow in section 5.

## 5 Reconstructing a sound change

The experimental sample considered in this paper represents a fairly small window into the variation that exists in Turkish; the range in apparent time is small, and the set of participants is fairly homogeneous in sociolinguistic terms. In much the same vein as in ??, therefore, the construction of any account of the trajectory of change and the historical context for innovation is necessarily reliant on other strands of evidence and argumentation.

### 5.1 Unexpected extensions

The case with which this paper concerns itself is properly introduced in section 2. The role of the preceding discussion was the (partial) delineation of the complicated relationship between synchronic phonology, sound change, and statistical typologies; in this section, we consider (with slightly more depth) the emergence of unnatural-seeming phonological activity as a consequence of sound change, and its relevance to the pathway by which grammatical generalisations arise.

What is at issue is the pathway by which ‘un-phonetic’ generalisations can emerge. If a phonological pattern originates in a speaker-listener mismatch over natural and well-grounded phonetic features, then the appearance of unnaturalness must arise from some implementational change subsequent to the original phonetic effect; when can such changes arise, and what scope can they have? We might see this as involving a move from phonetically-driven to phonologically-driven (that is, more abstract) control (Kiparsky, 1995; Janda, 2003): as a rule ‘dephoneticises’ and is less strictly coupled to the action of phonetics, the potential for generalisation to a set of environments in which its action is less physically and perceptually well-motivated exists. If this is the case, then one suggestion is that the *craziest* (Bach and Harms, 1972) rules are the most extreme members of a broader typology of the end-products of change, and that what we should expect to see is rules in *intermediate* historical states. We illustrate this point with the two sketches that follow; in both examples, an original phonetically-grounded state can be deduced, but phonologisation has driven the emergence of what we call a ‘natural class plus’, in which a pattern generalises to an ‘L-shaped’ set of contexts. The further question we will revisit throughout this paper concerns the extent to which we can adduce information about the pathway of phonologisation from *non-unnatural* cases: do the regular-seeming products of phonological change carry traces of the potential for more arbitrary generalisations?

### 5.1.1 Northeastern Swiss German

A particularly high degree of phonological variation exists in the German dialects of north-eastern Switzerland; phonological systems differ from ‘one village to the next in the same sub-dialect of a single canton’ (Keel, 1982). In the Swiss German varieties of the canton of Schaffhausen, an assumed historical rule lowering pre-rhotic (whether within the same syllable, or across a syllable boundary) o → ɔ has undergone generalisation into several distinct systems, summarised overall in example 12 (Keel, 1982; Janda and Joseph, 2003). All dialects show o → ɔ before r (example 11); variation arises when the domain of the rule is further extended. Evidence for further *synchronic* conditioning of this lowering rule comes from alternation under affixation example 14: plurals and diminutives of words with /ɔ/ have /ø/, under both the loss of sonorant-triggered lowering and a dialectally-widespread morphologically-conditioned fronting (Garrett, 2014).

*Example 11* Schaffhausen German: /o/-lowering in all dialects before /r/ (Keel, 1982):

børə	‘to bore’
hörn	‘horn’
kwɔrfə	‘thrown’

*Example 12* Schaffhausen German: dialectal variation (Keel, 1982):

[Schaffhausen proper:] o → ɔ before r and *nasals*; all above, and:

xɔ.məd	‘come-PL.IND.’
tɔnə	‘to drone’
fɾɔm	‘pious, good’

13 of the surrounding villages: o → ɔ before r, *nasals*, and *coronals*; above, and:

ʃtɔtsə	‘to push down’
rɔss	‘horse’
ʃnɔdərə	‘to stir’

[17 villages:] o → ɔ before r and *coronal obstruents*; but *not* nasals.

[5 villages:] o → ɔ before r, *coronal obstruents*, and *non-coronal obstruents* except b.

*Example 13* Schaffhausen German: /o/-lowering never applies before /l/ (Keel, 1982):

folk	‘full’
holə	‘to fetch’
bollə	‘bud’

*Example 14* Schaffhausen German: morphologically-conditioned loss of [ɔ] (Garrett, 2014):

/bɔdə/ → bødə	‘floor, ground’-PL
/lɔxxə/ → løxxli	‘hole’-DIM

The rhotic is the only trigger for o-lowering that persists exceptionlessly across (sub-)dialects; this is taken by Keel (1982); Janda and Joseph (2003); Mielke (2008) to be evidence that the rhotic is the original environment in which the o-lowering rule was first instantiated. In Schaffhausen proper, the environment for the rule is given by part of the natural class of sonorants: the consonantal sonorants absent the lateral /l/. In 17 villages of the Schaffhausen canton, the environment for the rule is given by the set {r, coronal obstruents}, in which coronality is shared but manner differs. In 13 other villages, these generalisations are superimposed: o-lowering applies before both nasals and coronal obstruents.

The extensions that arise in the different varieties do not lend themselves as well-behavedly to further categorical generalisations; nor do they strictly track the set of contexts in which a vowel-lowering rule is best-motivated. Some implicational generalisations are observed—non-coronals are involved only if coronals are involved—but not all cases admit further hierarchical generalisations: obstruents and nasals seem to be involved (or not) freely, subject to no further implicational requirement. It is the case in every instantiation of the o-rule that the rhotic must be involved, due to its status as the ‘original’ trigger; we may further adduce example 12 that  $\text{o} \rightarrow \text{o}$  applies before non-coronal obstruents iff it applies before coronal obstruents, but the coronal obstruents and nasals behave orthogonally here—either may be a triggering environment independent of the other.

Of course, the observation that the distribution of patterns in Schaffhausen German involves variable extension from a single plausible phonetic precursor is not a new one; this insight underlies the discussion of the case in Janda and Joseph (2003); Mielke (2008). Janda and Joseph use cases of this type as evidence for their ‘big bang’ theory of sound change, which operates along the lines of the sketch in example 15; in less specific terms, what is being described here is really a conception of *phonologisation* that is early and abrupt, rather than gradual.

*Example 15* The Big Bang theory of sound change (Janda and Joseph, 2003), and its relationship to other accounts.

1. All sound changes begin as very ‘small’, very localised, **purely phonetic** effects, over a **short temporal span**; the obligatory presence of early phonetic conditioning is after e.g. Ohala (1981).
2. The future trajectory of generalisation is determined by this original ‘burst’;
3. As an innovation generalises and spreads, phonetic conditions are supplanted; **phonological and/or sociolinguistic** conditions take precedence.
4. At the latest stages, reanalyses lead to new morphological or lexical conditioning; cf. Bermúdez-Otero (2007)’s ‘domain narrowing’.

The relationship of this schematic to examples 11, 12 and 14 is in the inherent variability of the patterns. If it must be the case that sound change has some initial phonetic trigger, then in order for the different dialects to show this variety of implicational-unrelated, phonetically-unjustified, but very different alternations, ‘control’ over the development of these patterns must be ceded to the phonology. What this argument does not, however, cover, is why we see these generalisations in particular; if we say that the phonology drives the formation of generalisations to novel environments,

then what we require in our theory of phonology is a mechanism by which these novel environments are selected or ruled out.

### 5.1.2 Georgian

In Georgian (Butskhrikidze, 2002, p. 90), the non-high vowels /ɑ/ and /e/ followed by coda {m, n, l, r, v} are deleted when -V(C) suffixes are appended<sup>23</sup>; it can be argued (*ibid.*) that /v/ in Georgian should be analysed as a sonorant and patterns generally with the set {m, n, l, r} across processes. The case of /o/ is more complicated. If followed by the labial /m/ or /b/, /o/ deletes; elsewhere, it alternates with v example 16.

*Example 16* Georgian: o-deletion and o-v alternation. (Butskhrikidze, 2002, p. 82 & 95)

/diyomi-is/	[diymis]	toponym-GEN
/sap'oni-is/	[sap'nis]	soap-GEN
/xoxob-is/	[xoxbis]	pheasant-GEN
/mindor-is/	[mindvris]	field-GEN
/p'amidor-is/	[p'amidvris]	tomato-GEN
/nigoz-is/	[nigvzis]	nut-GEN

*Example 17* Georgian: syncope in VCV(C) if intervening C is {m, n, l, r, v}. (Butskhrikidze, 2002; Butskhrikidze and van de Weijer, 2001)

/mercxal-is/	[mercxlis]	swallow-GEN
/t'omara-it/	[t'omrit]	sack-INST
/jvel-is/	[jvlis]	deer-GEN
/bal-eb-i/	[blebi]	cherry-PL-NOM
/xed-av-a/	[xedva]	see-THEM-INF
/se-i-p'χ'ar-ob/	[seip'χ'rob]	'you will arrest'
/ga-t'ær-i/	[gat'ri]	'you will cut'
/xar-av-a/	[xvra] <sup>24</sup>	gnaw-THEM-INF

*Example 18* Georgian: no syncope for non-V(C) affixes or non-sonorant finals.

/mercxal-ma/	[mercxalma]	swallow-ERG
/k'amat-is/	[k'amatis]	debate-DAT

If we admit /v/ as a sonorant in Georgian<sup>25</sup>, examples 17 and 18 then describe a pattern whose environment is given entirely by the feature [sonorant]: pre-sonorant vowels undergo syncope, pre-obstruent vowels do not. However, Butskhrikidze gives cases example 19 in which this pattern extends to non-high vowels followed by the voiced stop /b/:

<sup>23</sup> Except the nominative /-i/: this is appended to consonant-final stems to satisfy constraints on well-formedness, and does not trigger any alternation.

<sup>24</sup> Both the vowel in /-av/ and the root vowel are deleted; this shows a further case of v-metathesis, for details of which see Butskhrikidze and van de Weijer (2001).

<sup>25</sup> Which is supported more generally in the phonology by its behaviour in syllabification; see Butskhrikidze and van de Weijer (2001); Butskhrikidze (2002), again.

*Example 19* Georgian: syncope before /b/. (Butskhrikidze, 2002)

/k'ak'ab-is/	[k'ak'bis]	partridge-GEN
/xoxob-is/	[xoxbis]	pheasant-GEN

The distinction between pre-sonorant contexts and pre-obstruent contexts is not arbitrary with respect to the pattern itself; but a potential argument for the emergence of this alternation from sonority or syllacticity is not straightforward to reconcile with the optional participation of the obstruent /b/. The parallel we want to draw between the cases in Schaffhausen German and in Georgian, and eventually extend to the novel case in Turkish, is then as follows. Both the Schaffhausen /o/-lowering and the Georgian mid-vowel syncope apply in environments which are supersets of some ‘sensible’ set of environments, with respect to *both* phonetic grounding and natural class behaviour. /o/-lowering generalises to various L-shaped sets that don’t constitute natural classes in themselves, but that are supersets over multiple reasonable natural-class generalisations. Lowering a mid vowel has ample phonetic motivation pre-rhotic, but it is not straightforward to deduce that there exists equal phonetic motivation in various pre-obstruent positions to which the change optionally generalises—especially given the absence of any requirement on the moraic position of the trigger segment. Mid-vowel syncope in Georgian extends to pre-/b/ mid vowels, despite the violation of natural-class boundaries; as in the previous case, this has some intuitive (though currently poorly-defined) origin in the similarity of /b/ and of a peripheral member of the class of sonorants, /v/. Consider the inventory for Georgian given in table 15; the set of segments that are uncontroversially sonorant in Georgian is highlighted, as is the /b/.

p	t	ts	ts'	k			
p'	t'	ts'	ts'	k'			
		s		x	χ'	h	
b	d	dz	dz'	g			
v	z		ʒ	ɣ			
m	n						
	l						
	r						

**Table 15** The Georgian consonant inventory, adapted from Butskhrikidze (2002).

The ordering of rows in table 15 is somewhat (intentionally) impressionistic; the intent here is solely to present a sketch of where the syncope-triggering segments might fall in a notional ‘similarity space’. Much as in Schaffhausen German, it was difficult to claim that phonetic motivation for a change in vowel height existed in every environment, it’s here difficult to construct either a phonetic or a perceptual reason for syncope to apply before this set of segments. The nearest thing we then find to a

unifying generalisation is the *similarity* of /b/ to nearby members of the more sensible set; in particular, to /v/, which itself could plausibly be understood to represent the original trigger for syncope. Butskhrikidze (2002, p. 88) claims that C+v/ sequences are realisationally most like C<sup>w</sup>, single segments with secondary labialised articulation; see also the general special status of labial segments in the phonology, as example 16. If this is the case, then of all possible C+son sequences, C+v/ constitutes the best onset, and thus also the best possible output of V-deletion in CVC under resyllabification: despite the well-known tendency in Georgian toward large clusters, simple onsets are still preferred, where possible, throughout the phonology. It's then plausible that syncope giving rise to C+v/ (as in example 17: /xed-av-a/ → [xedva] or [xed<sup>w</sup>a]) was preferentially phonologised on this basis, and extended to the rest of the sonorants; no such property holds for underlying /b/, however.

I will argue that the Turkish case shows a similar lack of what we will call ‘first-order’ groundedness. Cases of this type must represent abstract generalisations, in which the role of *phonological* information is crucial; the operation of the phonological pattern shifts in scope from an area of good grounding to an area of poorer grounding, related to the first in some representational way. Note that this does not imply development *de novo*, with no phonetic precursor at all, but rather that the leap from the domain of gradient phonetics to the domain of categorical phonology arises early in the stabilisation of a rule, allowing biases towards particular phonetic precursors to dissipate.

## 5.2 Mid vowels in non-standard dialects and related languages

Although we have referred generally in this paper to ‘Turkish’ without further qualification, this should be understood to refer to the standard variety, very broadly (Lewis, 1967) understood to be the speech of the upper classes of Istanbul and Ankara. This is unsurprisingly reductive; there is substantial dialectal variation, particularly in phonology, across Turkish-speaking regions. Karahan (1996) has a *relatively* recent and thorough classificatory dialectology, postulating three major dialect groups in Anatolia: western and eastern Anatolian groups separated from each other roughly by the boundary of the Euphrates, and a north-eastern group encompassing the dialects of Trabzon, Rize, and areas surrounding (for which last a very comprehensive phonological account in English exists, Brendemoen 2002).

This subsection considers the systems in some of these non-standard varieties, to the extent that the limited data allow. We have access to very little conditioned experimental data for any regional varieties; conclusions drawn in sections 5.2.1 and 5.2.2 are based on what is often passing reference in the literature, and should therefore be assigned the appropriate degree of scepticism in consideration. Section 5.2.3 shows the vowel space for a single divergent speaker, M03, who was excluded from the analysis due both to gender and to a markedly distinct system; this speaker’s production suggests the presence of gradient but statistically non-negligible pre-rhotic vowel height movement, plausibly a phonetic precursor to the more fully phonologised effects seen both in the standard variety and in Trabzon Turkish.

### 5.2.1 Western Anatolian rhoticity loss

An often-cited minor example of compensatory lengthening triggered by syllable-final /r/-deletion (Sezer, 1986; Kavitskaya, 2002) in Western Anatolian Turkish dialects incidentally suggests that these dialects show additional rhotic-triggered vowel height effects, which persist even when the rhotic is deleted in the surface output. The data in example 20 are Sezer's transcription of Korkmaz (1965), and the pattern itself seems restricted to the environment of compensatory lengthening (that is, to syllables underlying closed by a surface-deleted rhotic).

*Example 20* Compensatory lengthening and mid-vowel lowering in Western Anatolian dialects. (Sezer, 1986, p. 241)

Standard Turkish	Western Anatolian	
var	va:	'there is'
verdi	væ:.di	's/he gave'
giderler	gi.dæ:.læ:	'they go'
pisirir	pɪsɪræ:	's/he cooks'
verir	viri:	's/he gives'

### 5.2.2 Sonorants and velars in Trabzon dialects

In the traditional 'East Anatolian' dialects, the distinction between /e/ and /æ/ is a phonemic one; this is also the case in Azerbaijani, for which more description is available (Dehghani, 2000; Schönig, 1998). In the dialects of Trabzon and the neighbouring regions, in the north-east of Anatolia, (Brendemoen, 2002, p. 53) describes an ongoing merger between phonemic /e/ and [i], which proceeds *unless* blocked by an immediately following /l, y, ɣ/. There is further free variation, if /e/-[i] alternation does not apply, between [e] and [æ] in pre-sonorant and pre-velar positions /r, l, k, y, ɣ, n/. An overview, with incomplete coverage of the environments in which Brendemoen claims an effect arises, appears in examples 21 and 22<sup>26</sup>; a full dataset is unavailable. 'Standard Turkish' given below is phonemic, and does not indicate /e/ → [æ] etc.

*Example 21* /e/-raising and [i]-blocking in Trabzon dialects.

Standard Turkish	Trabzon	
/erkek/	er.kek ~ er.kik	'male'
/köp/ <sup>27</sup>	kep ~ kip	'many'
/et/	et ~ it	'do/reach'
/kel/	kel *kil	'come'
/ejer/	ezer *ezir	'saddle'

*Example 22* /e/-lowering in Trabzon dialects.

<sup>26</sup> Brendemoen (2002) marks aspiration on some of these examples, but inconsistently; as it's unclear if this is contrastive, and unclear whether its absence in the text indicates true absence or simply lack of disambiguatory relevance, we have omitted this here.

Standard Turkish	Trabzon	
/geldi/	gæl.di	‘came’
/gid-er-ken/	gidærgæn	‘while going’
/benzer/	bænzer	‘similar’
/ben/	bæn	‘I’
/yemek/	yemæk	‘food’

Despite incomplete evidence, the best-guess deductions that we can make about the overall attested space of variation appear in table 16:

Dialect	Rule type	Triggers
Trabzon (NE Anatolian)	/e/-[i] blocked, /e/-[æ] promoted	/r, l, ɣ, ɳ/ block; /r, l, k, ɣ, ɳ, n/ cause
general Eastern Anatolian	/e/, /æ/ have phonemic status	
Western Anatolian	/e/-[æ]	/r/
‘Standard’ Turkish	/e/-[æ], /ø/-[œ]	/r, l, m, n/

Table 16 The state of mid-vowel rules across consensus Turkish dialects.

### 5.2.3 Divergent individuals and phonetic detail

In the data we present in section 4.1, which represent in total a fairly coherent variety of the language, a relationship holds exceptionlessly between sonorant environment and vowel height. For the variant systems in the ‘non-standard’ dialects that appear immediately above, this is also the case; these systems differ from the consensus variety in the details of their phonological conditioning, but shared consistently the presence of *phonetically abrupt* mid-vowel lowering. Consider the non-normalised F2×F1 space in figure 37 below; each point represents one token produced by a single speaker (table 1: M03) from the eastern city of Kars, and the whole is indicative of a very divergent dialect.

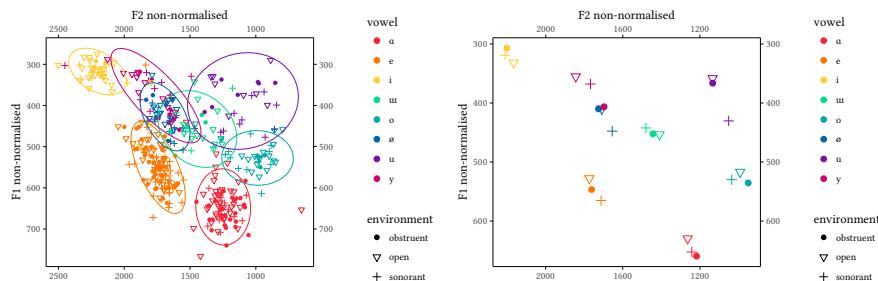


Fig. 37 Dialectal divergence: F2×F1 space (non-normalised) for speaker M03, Kars. Left: each point corresponds to a single token, measured at the midpoint of the vowel. Right: points represent the average across all tokens for each vowel in each environment.

<sup>27</sup> The Trabzon dialects already have a partial /e/-/ø/ merger; additionally, /køp/ is itself an archaism/regionalism.

Due to the clear difference in underlying system, and to the relative paucity of data (lexical items with which the speaker was entirely unfamiliar were omitted in recording, and thus only 158 of 220 wordlist items were available for analysis), this speaker is not considered in combined analyses above.

We might impute to this speaker's overall system, however, a status as one 'extreme' of the possible realisational space for Turkish /e/, across dialects: although some spread in possible F1 appears, with a 95% confidence range of [451, 637] Hz, this does not appear to be distributed strictly by *environment* as previously coded (as a factor with levels: obstruent, sonorant, no-coda), and simple linear models<sup>28</sup> suggest no statistically meaningful effect of environment. An alternative model in which F1 is a function of *rhoticity*, however,  $F1 \sim \text{rhotic}$ , where tokens are coded by the presence or absence of a following /r/ coda, gives a significant effect due to /r/, an F1 increase of  $[53.33 \pm 15.68]$  Hz,  $F(1,115) = 7.353$ ,  $p = 0.007721$ .

### 5.3 Variation (and theme)

The system in the Turkish of Western Anatolia, as given in section 5.2.1, is a comparatively simple one in its targeting and conditioning; lowering applies to the unrounded mid vowel if it precedes a rhotic. In both the standard variety that constitutes much of the focus of this paper, and in the Trabzon Turkish considered in section 5.2.2, mid vowel lowering is subject to more complex and extensive conditioning; while in the speech of the informant from Kars presented immediately above, little is seen other than a small phonetic effect pointing in a generally similar direction.

The single property that unites this disparate set of systems is the presence of an effect triggered by the rhotic. In the standard variety and plausibly in Trabzon Turkish, this effect has phonologised; that is, there now exists an effect whose control is within the domain of the phonological computation, rather than in the domain of mechanical phonetics. We cannot make a judgment as to the phonological status of vowel height in Western Anatolian Turkish, but we can note that the pre-rhotic effect therein is at least well-documented and understood to be phonetically discontinuous; in the speech of the informant from Kars, the balance of evidence suggests that the pre-rhotic effect is not (yet) a grammatical one.

The disunity across systems is then essentially emergent from phonologisation itself; in asking the question of what happened in Turkish, we are implicitly also asking why it is that the process of phonologisation produced this particular rule, with this particular domain, in the standard variety; a different set of environments in the Trabzon variety; and did not generalise beyond the rhotic in other regions of Anatolia. In the standard variety, we can take into account the apparent presence of a pre-/z/ effect (section 4.5), and describe the operation of mid-vowel lowering as involving the active class /r l n m z/; in Trabzon, the relevant environment is instead /r l n ɳ y/

<sup>28</sup> Generalised linear mixed-effects models were used above to capture the effects of any possible inter-speaker or item-specific variation, modeled with random slopes. For the case of a single-speaker subset of the data, accounting for speaker-specific variation is unnecessary; the models referenced here are linear regressions with one dependent variable, F1, and the categorical predictors *environment* and *stress* (and their interaction).

k/, with the exclusion of the labial nasal but with the inclusion of a phonemic dorsal nasal<sup>29</sup> and the dorsal obstruents. Not all these environments are created equal, from the point of view of the action of phonetics; as in section 4.2.2, the Turkish lateral is so heavily fronted in the context of underlying front vowels (whether or not those vowels are targeted on the surface by the phonology) that it should moderately disfavour lowering from an articulatory and acoustic perspective.

There are several explanatory problems: why are the individual mid vowels not equally targeted, how do triggering classes of this type arise, and what is it that drives the differences in final phonologised state between standard Turkish and Trabzon Turkish? Both varieties show an active class that mixes sonorants and obstruents; in the standard variety, all sonorants are involved, while Trabzon omits the labial, but in both cases the phonetically-unfavourable lateral participates exceptionlessly. The essential conjecture that we will propose here is that this set of facts is evidence for a neat diachronic pathway, which in its structure essentially recapitulates the logic of ???. The persistence of the rhotic across varieties and across types of phenomenon (both phonetic and phonological) lends itself to the idea that the rhotic constitutes an initial phonetic precursor to the phonological change that we observe; as we proposed in section 5.2.3, it appears to have a phonetic effect in varieties of Turkish that do not in themselves show any sign of categorical /e/-effects, and . If we then posit that the phonologisation of lowering in the mid vowels in Turkish is driven by an initial, functionally-grounded and well-motivated effect of a rhotic on a preceding vowel; the subsequent decision process contained in stabilisation and generalisation of the initial rule extends over the set of mid vowel undergoers and the full consonantal inventory of the language, proceeding according to similarity to the initial trigger. The difference in resultant activity between the standard variety and the variety of Trabzon we might then attribute to differences in the consensus computation of similarity: for speakers of the Trabzon variety, similarity of the rhotic precursor to the dorsal nasal and thereby to dorsal obstruents<sup>30</sup> takes precedence over similarity to the labial nasal. This process of generalisation is fundamentally *abstract*, in the sense that after the initial actuation it pays relatively little ‘attention’ to phonetic substance; this allows generalisation both to the lateral, and eventually to the voiced obstruent /z/.

### 5.3.1 On diachrony, again

How did the sound system of Turkish get into this state? Recall the examples from Schaffhausen German and from Georgian given in section 5.1.1 and section 5.1.2 respectively. What these cases had in common was the L-shaped nature of the classes involved; recall the active class given in table 15 for Georgian, and the inventories and active classes given below in table 17 for the micro-variation across the local varieties of Schaffhausen German.

---

<sup>29</sup> Marginal to the standard variety.

<sup>30</sup> Which are better promoters of lowering and backing than many of the other segments involved; there is then an independent question of whether a second phonetic influence is to be considered in the Trabzon variety.

$p^h$	$t^h$	$k^h$	
b	d	g	
pf	ts	kx	
f	s	ʃ	x
z	ʒ		h
r			
m	n	ŋ	
l			
j			

$p^h$	$t^h$	$k^h$	
b	d	g	
pf	ts	kx	
f	s	ʃ	x
z	ʒ		h
r			
m	n	ŋ	
l			
j			

$p^h$	$t^h$	$k^h$	
b	d	g	
pf	ts	kx	
f	s	ʃ	x
z	ʒ		h
r			
m	n	ŋ	
l			
j			

$p^h$	$t^h$	$k^h$	
b	d	g	
pf	ts	kx	
f	s	ʃ	x
z	ʒ		h
r			
m	n	ŋ	
l			
j			

**Table 17** Consonant inventories and active classes in Schaffhausen German, adapted from Mielke (2004). The classes of alternation-triggering segments across the different varieties have been highlighted in each case.

Both the pattern in Georgian, in which deletions applied before a coherent set of sonorants and the voiced labial stop /b/, and the rather wide set of variants depicted above in table 17 are alike, in that we can think about them as containing a *core* that represents an original actuator. In Georgian, this is likely to be the /v/, triggering alternations first in the labial /o/, then generalising to the other non-high vowels and to the sonorants, and to the /b/ via shared labiality. In Schaffhausen German, this is likely to be the rhotic; the geography of variation in this case shows us that the rhotic is the only *consistent* participator across the environments in which the rule of /o/-lowering is instantiated. Extension from the rhotic then proceeds in the direction either of the ‘nearby’ sonorants, or of the coronal and then non-coronal obstruents. In both cases, then, the ultimate determiner of the set of triggers (and the set of undergoers) is the action of diachrony.

Recall the analysis we proposed for the set of undergoers of sonority-driven alternation in ??: in that case, the problem was not any individual set of undergoers in any individual language, but rather the disparity between the relative frequency with which each sonorant segment acted as an undergoer in the overall typology, and the predictions of synchronic accounts of the actual alternations. The solution was diachronic; the alternations themselves were and are synchronically-conditioned, but the composition of the set of undergoing segments is set by the deterministic pathway a given language’s phonology takes through the space of possible rule-extensions and rule-generalisations, rather than by an explicit synchronic principle. At the same time, our claim is that there *is* a ‘synchronic’, or at least grammatically-encoded principle

that constrains that deterministic pathway: the probability that a new segment is selected as an undergoer is given by its similarity to the set of segments that are already included in the statement of the rule.

Is this a style of analysis that we can extend to the Turkish case? Consider the description in table 18 of the case in standard Turkish that we establish in this paper, and of the variation in the Trabzon dialects given by Brendemoen (2002).

'Standard' Turkish					Trabzon Turkish				
f	s	ʃ		h	f	s	ʃ		h
v	z	ʒ			v	z	ʒ		
	r					r			
m	n				m	n		ɳ	
b	d	ðʒ	g		b	d	ðʒ	ɣ	
p	t	ʈʃ	k		p	t	ʈʃ		k

**Table 18** Consonant inventories and active classes promoting the rule of /e/-lowering in two varieties of Turkish; the standard Turkish of this paper, and the Trabzon dialect given by Brendemoen (2002).

There is a clear analogy between the environment for /e/-lowering in both the standard and the Trabzon variety, the presence of phonetic pre-rhotic lowering in the speech of the informant from Kars, and the descriptions we assign to the L-shaped generalisations in Swiss German and in Georgian. In the Trabzon variety, the extension of phonetically-driven pre-rhotic lowering from its original environment to the final set takes a different path through the inventory from the case that we cover in this paper, bounded more closely by considerations of place; generalisation to coronals and dorsals is strongly preferred over generalisation to the labials. The crucial property that all these cases share is that they do not straightforwardly admit expression as a more traditional featurally-based class; we cannot really establish a simple conjunction of features that does not then include unwanted segments. At the same time, they are far from being the truly ‘crazy’ rules of Bach and Harms (1972); there is a clear logic of transitivity by which they can be derived sequentially from the original actuator.

What is the relationship here to similarity? our argument, as throughout, is that it is desirable to be able to understand convex class definitions of this type as deriving from the action of an internally-coherent process of phonologisation; and if we understand this to be the case, then we demand a theory of similarity. The reason to attribute organisation of this type to similarity rather than to another organising principle lies in the need to predict both likelihood and optionality; this calls for a parameter that behaves more like a continuous distance or like a transition probability than a strictly hierarchical principle.

## 6 In sum

The first contribution of this paper is the descriptive one; that is, the establishment of the existence, categoricity, and interest of the problem of mid-vowel lowering in Turkish. We have demonstrated throughout (section 4) that the Turkish mid vowels are subject to alternations conditioned by [ $\pm$ sonorant] coda and by /z/ (and sensitive explicitly to the syllable boundary and not simply the presence or absence of the segment; figure 2); that the alternation in /e/ displays a much more discontinuous and categorical-seeming distribution in the phonetic parameter space (section 4.1), is subject to a larger set of exceptions (section 4.4), and *lacks* the apparent dependence on triggering segment seen for /ø/ (section 4.2) and for dialectally-divergent speakers from whose systems phonological /e/-lowering is definitively absent (section 5.2.3). We have suggested further that the ‘initial’ state of the Turkish mid vowel system most closely resembled the synchronic state of unstressed /ø/, in which a process of raising in unstressed open syllables interacts with phonetically-driven, gradient lowering triggered by the rhotic; the most innovative state resembles the overall pattern in stressed /e/.

More central to the issues that we raise in ?? are the following remarks. First, generalisation of /e/-lowering *at least* to the negative aorist /-mez/, outside the [sonorant] class previously characterised, is well-underway, and is discontinuous rather than gradual with respect to manipulation of the phonetic parameter involved. Second, the phonetic properties of the Turkish lateral suggest that it should exert little to no coarticulatory pressure to *lower* a vowel, but there is no statistically-significant difference between pre-lateral tokens and pre-rhotic tokens, despite the rhotic’s significantly better inherent properties as a precursor to this type of sound change. The net implication of the descriptive problem is therefore a problem of class formation and structure: what is it that causes a phonological rule to involve a set of segments that do not conform either to representational unity or phonetic unity?

From an explanatory point of view, we argue that this disparate collection of facts allows us to derive a well-ordered set of diachronic hypotheses, despite the poor time-depth of the experimental sample; the intent of this claim is to illustrate the potential relationship between the selection of active classes and the trajectory along which phonological change progresses, a problem with which we are also concerned in ???. Differences in categoricity and continuity, dependence on trigger, and sensitivity to lexical and prosodic effects suggest that /ø/ is *behind* /e/ in the process of rule-generalisation; the persistence of small-scale effects in the rhotic, the relevance of the rhotic to the state of rules in non-standard varieties, and the existence of phonetic effects targeting pre-rhotic vowels in varieties that show no categorical phonological rule suggests to us a diachronic pathway involving a series of successive generalisations from a functionally-motivated rhotic precursor.

**Acknowledgements** Redacted for review.

### Conflict of interest

The authors declare that they have no conflict of interest.

### References

- Adank P, Smits R, Van Hout R (2004) A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5):3099–3107 13
- Árnason K (1999) *The Phonology of Icelandic and Faroese*. Oxford University Press, Oxford 9
- Bach E, Harms RT (1972) How do languages get crazy rules? In: Stockwell RP, Macaulay RKS (eds) *Linguistic change and generative theory*, Indiana University Press, Bloomington, pp 1–21 47, 58
- Barnes J (2007) Phonetics and phonology in Russian unstressed vowel reduction: a study in hyperarticulation, mS., Boston University 38
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48 18, 19
- Beddor PS (1993) The perception of nasal vowels. In: Huffman MK, Krakow RA (eds) *Nasals, Nasalization, and the Velum*, *Phonetics and Phonology*, vol 5, Academic Press, San Diego, pp 171 – 196 10
- Beddor PS, Krakow RA, Goldstein LM (1986) Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook* 3:197–217 10
- Bermúdez-Otero R (2007) Diachronic phonology. In: de Lacy P (ed) *The Cambridge Handbook of Phonology*, Cambridge University Press, Cambridge, p 497–517 49
- Bermúdez-Otero R (2015) Amphichronic explanation and the life cycle of phonological processes. In: Honeybone P, Salmons J (eds) *The Oxford handbook of historical phonology*, Oxford University Press, Oxford, pp 374–399 44
- Bilgin O (2016) Frequency effects in the processing of morphologically complex turkish words. Master's thesis, Bogaziçi University, URL [http://st2.zargan.com/public/resources/turkish/frequency\\_effects\\_in\\_turkish.pdf](http://st2.zargan.com/public/resources/turkish/frequency_effects_in_turkish.pdf) 46
- Blaskovics J (1964) *Gagavuz Türkçesi Grameri*. Türk Dil Kurultayında Okunan Bilimsel Bildiriler, Türk Tarih Kurumu Basımevi, Ankara 4
- Blust R (2013) The Austronesian languages. *Asia-Pacific Linguistics*, Canberra 7, 8
- Boersma P, Weenink D (2017) Praat: doing phonetics by computer. URL <http://www.praat.org/> 12
- Botma B, Van Oostendorp M (2012) A propos of the Dutch vowel system 21 years on, 22 years on. In: Botma B, Noske R (eds) *Phonological explorations: Empirical, theoretical and diachronic issues*, Mouton de Gruyter, Berlin/Boston, pp 1–16 7
- Bradley TG (2010) Mid front vowel lowering before rhotics in Ibero-Romance. In: Herschensohn J (ed) *Romance Linguistics 2010*, John Benjamins, Amsterdam, pp 63–78 9
- Brendemoen B (2002) *The Turkish dialects of Trabzon: their phonology and historical development*, vol 1. Otto Harrassowitz Verlag, Wiesbaden 52, 53, 58

- Butskhrikidze M (2002) The consonant phonotactics of Georgian, LOT Dissertation Series, vol 63. LOT, Utrecht 50, 51, 52
- Butskhrikidze M, van de Weijer J (2001) On v-metathesis in Modern Georgian. In: Hume E, Smith N, van de Weijer J (eds) Surface Syllable Structure and Segment Sequencing, Holland Institute of Generative Linguistics, Leiden University, pp 91–101 50
- Carter P (2002) Structured variation in British English liquids. PhD thesis, University of York 29
- Carter P, Local J (2007) F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association* 37(02):183 29
- Clements GN, Sezer E (1982) Vowel and Consonant Disharmony in Turkish. In: Hulst Hvd, Smith N (eds) The Structure of Phonological Representations, Foris, Dordrecht, pp 213–55 3
- Comrie B (1997) Turkish phonology. In: Kaye AS (ed) Phonologies of Asia and Africa (including the Caucasus), vol 1, Eisenbrauns, Winona Lake, Indiana, pp 883–898 4
- Dehghani Y (2000) A grammar of Iranian Azari: Including comparisons with Persian. Lincom Europa, München 53
- Féry C (2003) Markedness, faithfulness, vowel quality and syllable structure in French. *Journal of French Language Studies* 13(2):247–280 7
- Fruehwald J (2016) The early influence of phonology on a phonetic change. *Language* 92(2):376–410 19, 23
- Gallagher G (2016) Vowel height allophony and dorsal place contrasts in Cochabamba Quechua. *Phonetica* 73(2):101–119 7
- Garrett A (2014) Sound change. In: Claire Bowern BE (ed) The Routledge handbook of historical linguistics, Routledge, London & New York, pp 227–248 48
- Gussenhoven C (2007) A vowel height split explained: Compensatory listening and speaker control. *Papers in Laboratory Phonology* 9: Change in Phonology pp 145–172 8
- Göksel A, Kerslake C (2005) Turkish: A Comprehensive Grammar. Routledge, London 4, 6
- Hulst Hvd, Weijer Jvd (1991) Topics in Turkish Phonology. In: Boeschoten H, Verhoeven L (eds) Turkish Linguistics Today, E. J. Brill, Leiden, pp 11–5 3
- Inkelas S (1999) Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. In: Kager R, van der Hulst H, Zonneveld W (eds) The prosody-morphology interface, Cambridge University Press, pp 134–187 12
- Inkelas S, Orgun CO (2003) Turkish stress: a review. *Phonology* 20(01):139–161 12
- Iosad P (2012) Vowel reduction in Russian: no phonetics in phonology. *Journal of Linguistics* pp 521–571 38
- Janda RD (2003) Phonologization as the start of dephoneticization—or, on sound change and its aftermath: Of extension, generalization, lexicalization, and morphologization. In: Joseph BR, Janda RD (eds) The handbook of historical linguistics, Blackwell, Oxford, pp 402–422 47
- Janda RD, Joseph BD (2003) Reconsidering the canons of sound-change: Towards a “Big Bang” theory. In: Blake BJ, Burridge K (eds) Historical Linguistics 2001: Selected Papers from the 15<sup>th</sup> International Conference on Historical Linguistics,

- John Benjamins, Amsterdam, pp 206–19 48, 49
- Kabak B (2011) Turkish Vowel Harmony. In: van Oostendorp M, Ewen CJ, Hume E, Rice K (eds) *The Blackwell Companion to Phonology*, Wiley-Blackwell, Oxford, pp 2831–53 3
- Kabak B, Vogel I (2001) The phonological word and stress assignment in Turkish. *Phonology* 18(03):315–360 12
- Kahle D, Wickham H (2013) ggmap: Spatial visualization with ggplot2. *The R Journal* 5(1):144–161 11
- Karahan L (1996) *Anadolu Ağızlarının Sınıflandırılması*, Türk Dil Kurumu, vol 630. Atatürk Kültür, Dil ve Tarih Yüksek Kurumu, Ankara 52
- Kavitskaya D (2002) Compensatory Lengthening: Phonetics, Phonology, Diachrony. Outstanding Dissertations in Linguistics, Routledge, London & New York 53
- Keel WD (1982) Atomic phonology and phonological variation, vol 16. Gunter Narr Tübingen 48, 49
- Kewley-Port D, Watson CS (1994) Formant-frequency discrimination for isolated English vowels. *The Journal of the Acoustical Society of America* 95(1):485–496, DOI 10.1121/1.410024 22
- Kewley-Port D, Bohn OS, Nishi K (2005) The influence of different native language systems on vowel discrimination and identification. *Acoustical Society of America Journal* 117:2399–2399, DOI 10.1121/1.4785992 22
- Kiparsky P (1995) The phonological basis of sound change. In: Goldsmith J (ed) *The handbook of phonological theory*, Blackwell, Oxford, pp 640–670 47
- Kopkallı H (1993) A phonetic and phonological analysis of final devoicing in Turkish. PhD thesis, University of Michigan 4
- Korkmaz Z (1965) *Güney-Bati Anadolu Ağızları: ses bilgisi (phonetik)*. Ankara Üniversitesi DTCF Yayını, Ankara 53
- Kornfilt J (1997) *Turkish*. Routledge, London 6
- Krakow RA, Beddor PS, Goldstein LM, Fowler CA (1988) Coarticulatory influences on the perceived height of nasal vowels. *J Acoust Soc Am* 83(3):1146–1158 10
- Ladefoged P (2003) *Phonetic data analysis*. Blackwell, Oxford 9
- Lehiste I (1970) *Suprasegmentals*. MIT Press, Cambridge, MA 8, 32
- Lewis GL (1967) *Turkish Grammar*. Oxford University Press, Oxford 4, 6, 52
- Lindau M (1985) The story of /r/. In: Fromkin V (ed) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Academic Press, Orlando, pp 157–168 9
- Lindblom B (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35:1773–1781 8
- Lisker L (1974) On “explaining” vowel duration variation, tech. Rep. SR-37/38, Haskins Laboratories, New Haven, Connecticut 32
- Lobanov BM (1971) Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49(2B):606–608 13
- Maddieson I (1985) Phonetic cues to syllabification. In: Fromkin V (ed) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Academic Press, Orlando 7
- Maddieson I (1997) Phonetic universals. In: Hardcastle W (ed) *The Handbook of Phonetic Sciences*, Blackwell, Oxford, pp 190–232 8
- McCloy DR (2016) PhonR: tools for phoneticians and phonologists. R package version 1.0-7 13

- Mielke J (2004) The emergence of distinctive features. PhD thesis, The Ohio State University 57
- Mielke J (2008) The emergence of distinctive features. Oxford University Press 49
- Nevins A, Vaux B (2003) Consonant harmony in Karaim. In: Csirmaz A, Lee Y, Walter MA (eds) Proceedings of the Workshop on Altaic in Formal Linguistics, MIT Press, Cambridge, MA, pp 175–194 3
- Nichols J (2011) Ingush Grammar, University of California Publications in Linguistics, vol 143. University of California Press, Berkeley 7
- Nordhoff S (2009) A grammar of upcountry Sri Lanka Malay. PhD thesis, Utrecht University 6
- Ohala JJ (1981) The listener as the source of sound change. *CLS* 17:178–203 23, 49
- Ohala M, Ohala JJ (1992) Phonetic universals and Hindi segment duration. In: Second International Conference on Spoken Language Processing, Banff 8
- Peterson GE, Lehiste I (1960) Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32(6):693–703 8
- Proctor MI (2009) Gestural characterization of a phonological class: The liquids. PhD thesis, Yale University 9, 29
- R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/> 13
- Recasens D (2002) On the production characteristics of apico-alveolar taps and trills. *Journal of Phonetics* 19:267–280 9
- Recasens D (2012) Coarticulation in Catalan dark [h] and the alveolar trill: General implications for sound change. *Language and Speech* 56(1):45–68 29
- Recasens D (2014) Coarticulation and Sound Change in Romance. John Benjamins, Amsterdam 29
- Recasens D, Espinosa A (2005) Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: evidence from two Catalan dialects. *Journal of the International Phonetic Association* 35(1):1–25 29
- Recasens D, Pallarès MD (1999) A study of /j/ and /r/ in the light of the “dac” coarticulation model. *Journal of Phonetics* 27(2):143–169 9
- Riad T (2014) The phonology of Swedish. Oxford University Press, New York 9
- Rudis B, Bolker B, Schulz J (2017) ggalt: Extra Coordinate Systems, ‘Geoms’, Statistical Transformations, Scales and Fonts for ‘ggplot2’. R package version 0.4.0 11
- Sak H, Güngör T, Saracclar M (2008) Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In: Nordström B, Ranta A (eds) Advances in Natural Language Processing, LNAI, vol 5221, pp 417–427 46
- Schilling MF, Watkins AE, Watkins W (2002) Is human height bimodal? *The American Statistician* 56(3):223–229 14
- Schönig C (1998) Azerbaijani. Routledge, London & New York, pp 248–260 53
- Scobbie JM (2005) The phonetics phonology overlap. QMU Speech Science Research Centre Working Papers, WP-1 14
- Sezer E (1986) An autosegmental analysis of compensatory lengthening in Turkish. In: Wetzel L, Sezer E (eds) Studies in compensatory lengthening, Foris, Dordrecht, pp 227–250 53

- Silverman BW (1986) Density estimation for statistics and data analysis, vol 26. CRC press 14
- Solé MJ (2002) Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30(4):655 – 688 9
- Solé MJ, Ohala J (2010) What is and what is not under the control of the speaker: Intrinsic vowel duration. *Papers in Laboratory Phonology* 10 pp 607–655 32, 38
- Sproat R, Fujimura O (1993) Allophonic variation in English /l/and its implications for phonetic implementation. *Journal of Phonetics* 21(3):291–311 29
- Storme B (2017a) Contrast enhancement motivates closed-syllable laxing and open-syllable tensing, unpublished manuscript, lingbuzz/003700 8
- Storme B (2017b) The *loi de position* and the acoustics of French mid vowels. *Glossa: a journal of general linguistics* 2(1):64 8, 9
- Storme B (2017c) Perceptual sources for closed-syllable vowel laxing and derived environment effects. PhD thesis, Massachusetts Institute of Technology 6, 7, 8
- Strycharczuk P, Scobbie JM (2016) Gradual or abrupt? the phonetic path to morphologisation. *Journal of Phonetics* 59:76–91 29
- Tauberer J, Evanini K (2009) Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. In: Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009) 32, 38
- Toivonen I, Blumenfeld L, Gormley A, Hoiting L, Logan J, Ramlakhan N, Stone A (2015) Vowel height and duration. In: Steindl U, Borer T, Fang H, Pardo AG, Guekguezian P, Hsu B, O'Hara C, Ouyang IC (eds) *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, Cascadilla Proceedings Project, Somerville, MA, pp 64–71 8, 32, 38
- Turton D (2014) Variation in English /l/: Synchronic reflections of the life cycle of phonological processes. PhD thesis, University of Manchester 29
- Vigário M (2002) The Prosodic Word in European Portuguese. De Gruyter Mouton, Berlin, Boston 10
- West P (1999) Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics* 27(4):405–426 9
- Westbury JR, Keating PA (1980) Central representation of vowel duration. *Journal of the Acoustical Society of America* 67(S37) 8
- Yavuz H, Balci A (2011) Turkish Phonology and Morphology (*Türkçe Ses ve Biçim Bilgisi*). Anadolu Üniversitesi, Eskişehir 4