

## **A comprehensive model of phonological variation: grammatical and non-grammatical factors in variable nasal place assimilation**

*Abstract.* The past two decades have seen the development of several constraint-based models of phonological grammar that can handle variable phenomena. Most of these models, however, are purely grammatical and do not allow for the contribution of non-grammatical factors towards determining the frequency structure of variation. This paper reviews different approaches to phonological variation, focusing on how grammatical and non-grammatical factors codetermine patterns of variation. Based on this review a model is developed that incorporates influences from both grammatical and non-grammatical factors. The proposed model is grammar dominant in the sense that grammar defines the space of possible variation while non-grammatical factors only contribute towards the frequency with which the grammar determined forms are observed. Following Coetzee and Kawahara (2013), the model is developed in a version of noisy Harmonic Grammar that allows non-grammatical factors to scale the weights of constraints up or down. Two experiments on the perception of variable cross-word nasal place assimilation in English are reported, showing that this phenomenon is influenced by grammatical factors (place of articulation) and non-grammatical factors (speech rate and perceptual inhibition). An account for the results of these experiments is then developed in the noisy Harmonic Grammar model developed first by Coetzee and Kawahara (2013).

### Table of Contents

1. The place of variation in phonological theory .....	2
2. Cross-word nasal place assimilation in English .....	5
2.1. Production patterns of cross-word nasal place assimilation .....	5
2.2. Perception of cross-word place assimilated nasals .....	6
3. Experiments .....	8
3.1. Experiment 1: <i>n</i> ? .....	8
3.2. Experiment 2: <i>n/m/ng</i> ? .....	11
3.3. Comparing Experiments 1 and 2 .....	14
4. A noisy Harmonic Grammar account .....	15
4.1. The data .....	15
4.2. The basic grammar .....	16
4.3. Scaling the grammar .....	19
5. Concluding discussion .....	22
5.1. Grammar dominance .....	23
5.2. More on scaling factors .....	25
5.3. Final remark .....	26

## 1. The place of variation in phonological theory

Since the advent of generative phonology, the prospects of variation have changed drastically. Things developed from where variation was barely acknowledged as a phenomenon to the current situation where the adequacy of any phonological theory is partially determined by its ability to account for variation. This section provides a brief overview of some of the most important ways in which phonological variation has been approached over the past 40 years, focusing on two specific issues on which different approaches have diverged: (i) Does grammar only specify the possible variants, or does it contribute directly to the frequency structure of variation? (ii) How do grammatical and non-grammatical factors interact in phonological variation? See Coetzee and Pater (2011) and Coetzee and Kawahara (2013: section 1) for a more detailed review. After reviewing how models of phonological grammar have answered these questions over the past several decades, an alternative model is proposed that builds on these earlier approaches. In the model proposed here, it will be assumed that (i) grammar specifies what the possible variants are and contributes to the frequency structure of variation; (ii) non-grammatical factors contribute to the frequency structure of variation, but cannot add to the possible variants or invert frequency relations between variants specified by the grammar. The model proposed here can therefore be described as a “grammar dominant” model (more on this property in section 5.1).

Although variation did not feature prominently in the early years of the generative era, variation is consistent with the generative conceptualization of grammar (Kay and McDaniel 1979:152-154). Chomsky and Halle (1968:294), in fact, acknowledge variation as a part of the design of the grammatical model that they develop in *Sound Pattern of English*: “It is not necessarily the case that each deep structure determines a single phonetic representation; if the grammar contains optional rules or analyses, a given deep structure can underlie two or more phonetic transcriptions.” (see also Postal 1968:14-15.) What is not included in classic generative grammar’s notion of variability is information about the probability of application of an optional rule. As Kay and McDaniel state, “from the generative point of view ... grammar deals only with types, never with tokens” (1979:153). Token frequency was viewed as influenced by things such as memory limitations, the social setting of communication, real world knowledge, etc., and hence as being in the domain of “performance” rather than linguistics (Chomsky 1965:3-4, 10-15; see Newmeyer 2003:695-698 for a more recent but similar opinion).

Concurrently with the development of mainstream generative phonology, the Labovian variationist approach was also flourishing in the last three decades of the 20<sup>th</sup> century. Labov introduced the notion of the “variable rule” in 1969 (Labov 1969)—just one year after the publication of *SPE*. A Labovian variable rule differs from an optional *SPE* rule in two important ways: First, it includes information about the rate of application of the variable rule (i.e. it speaks to token frequency and doesn’t only specify the set of grammatical types). Secondly, it explicitly allows for interaction of linguistic factors with non-linguistic factors—variable rules can make reference to the communicative context, biographic information about the speaker, etc. Variable rules were given a mathematical interpretation by Cedergren and Sankoff (1974). In their approach, and later related developments (Paollilo 2002; Sankoff *et al.* 2005; etc.), variable rules were implemented as logistic regressions.

Two things should be noted about this implementation of variable rules: First, the factors entered into the analysis could be grammatical (place of articulation, stress, etc.). The contribution of such grammatical factors to the application rate of a variable rule is therefore directly estimated, so that grammar makes quantitative contributions to variation (contrary to *SPE* optional rules). Secondly, there is no formal difference in the treatment of grammatical and non-grammatical factors. The implication is that grammar does not occupy a privileged position relative to non-grammatical factors, and could in principle be trumped by non-grammatical factors. The reason for grammar’s non-privileged position is that variable rules were implemented as regression models, which are just statistical descriptions of

patterns in the data. These models are not restricted in a principled manner by the substance of linguistic theory, raising a question about their ontological status (Mendoza-Denton *et al.* 2003:107-110; Temple 2009:163). Is it reasonable to assume that “they are more than the constructions of the analyst, they are properties of the language itself” (Labov 1972:259)? Or should they be viewed as statistical descriptions of patterns in the data but not as models of the cognitive systems that produced the data (Fasold 1991; Kay and McDaniel 1979:152; Sankoff 1987:984; Walker 2012:398)?

With increased availability of large speech corpora, there has also been more research that explores these corpora for evidence of phonological variation, and that model the observed patterns statistically (Dilley and Pitt 2007; Raymond *et al.* 2006; Sonderegger 2012; etc.). Though not all of this research could be classified as being part of the Labovian variationist program, it shares with this program that data are usually described using statistical regression models. Like the variationist tradition, this tradition also tacitly assumes that grammatical factors contribute to the frequency structure of variation, and does not formally distinguish between grammatical and non-grammatical factors.

Although variation did not feature prominently in early main stream generative approaches, recent years have seen the development of several versions of generative phonological grammar intended specifically to deal with variation. Most of these were developed in constraint-based grammar, be that classic Optimality Theory (OT) (Anttila 1997, 2002, 2006; Anttila *et al.* 2008; Bane 2011; Coetzee 2004, 2006, 2009b; Reynolds 1994), stochastic OT (Boersma 1997; Boersma and Hayes 2001), noisy Harmonic Grammar (HG) (Coetzee 2009a; Coetzee and Kawahara 2013; Coetzee and Pater 2011; Jesney 2007), or MaxEnt models (Hayes and Wilson 2008). These models combine properties of both earlier generative approaches and of the Labovian variable rule approach. Like earlier generative approaches, they are primarily grammatical and pay little attention to non-grammatical factors. However, unlike in these earlier approaches, grammar specifies the frequency with which variants are observed in addition to the set of possible variants. Unlike the variable rule approach, however, these models are nearly all exclusively grammatical, and make no allowance for non-grammatical factors to contribute to variation.

Decades of research in both variationist sociolinguistics and the fields of speech perception/production have established that phonological variation is influenced by many factors of which grammar is only one. In fact, Bayley (2002:118) identifies “the principle of multiple causes” as a core principle that should direct the study of variation. Any formal model that purports to be a realistic model of the cognitive system underlying phonological variation therefore has to allow for both grammatical and non-grammatical factors to codetermine the patterns of variation. In order to develop a responsible model of phonological variation, it is therefore necessary to know how these two kinds of factors co-determine phonological variation. Are they independent from each other, making additive contributions to the frequency structure of variation? Or do they interact in more complex ways? Are they equal such either can trump the other? Or is one of the two kinds of factors dominant? Specifically, does grammar set strict limits of what counts as possible grammatical variants (as is assumed in the mainstream generative approaches)? Or can non-grammatical factors overrule the grammar-defined set of grammatical form (as is assumed in the variable rule approach)?

A general result of the variationist sociolinguistic tradition is that non-grammatical and grammatical factors are typically independent from each other (see, for instance, Bayley 2002:120; Guy 1991:5). Though members of the same speech community that occupy different positions on the socioeconomic ladder may differ in their absolute rate of application of a variable rule, the effect of grammatical factors on the application of the rule is usually the same for all members of the speech community. Labov’s New York City *r*-deletion study can serve as an example of this independence (Labov 1966:240). Although he found that speakers differed in how likely they are to produce pre-consonantal (*fourth*) and word-final *r* (*floor*) depending on their social class, *r* was more likely to be produced in word-final than pre-consonantal position for speakers from all social classes. Given this result, it will be assumed in the current paper that,

contra the variable rule tradition, grammatical and non-grammatical factors should be treated differently in a model of phonological variation.

Research in the variationist tradition has not focused on the question of grammar dominance—likely because no formal distinction is made between grammatical and non-grammatical factors in variable rules so that a dominance relationship between these two kinds of factors could not be expressed in this framework. The more conservative assumption, inline with how variation has been treated in mainstream generative approaches, is for grammar dominance. Since the variation patterns that have been described in the variationist research tradition are also consistent with an assumption of grammar dominance, the model used in this paper will assume grammar dominance, in agreement with earlier generative approaches and contra the variable rule tradition. This assumption of grammar dominance, however, should be taken as a strong hypothesis that is deserving of further research. Should further research find evidence that non-grammatical factors can override grammar, this aspect of the model developed here would need to be reconsidered. See section 5.1 for a detailed discussion of grammar dominance.

Though the generative models of variation developed over the past two decades did not include non-grammatical factors, the contribution of such factors were acknowledged, and ways of incorporating them were suggested. Van Oostendorp (1997), for instance, suggests that faithfulness constraints can be moved higher up in an OT grammar to account for more formal, and lower down for less formal speech registers. Boersma and Hayes (2001: Appendix C) similarly suggest that stylistic variation can be accounted for by ranking constraints differently for different stylistic levels. Coetzee (2009b), however, represents the first attempt at developing these suggestions into a fully operational model of phonological variation. In his model, developed in OT, faithfulness constraints move up or down in the hierarchy when evaluating words of different frequency. Coetzee (2009a, 2012) and Coetzee and Kawahara (2013) developed Coetzee’s original OT-based model in more detail in noisy HG.

In the Coetzee/Kawahara model, grammatical and non-grammatical factors both contribute to variation, but they do so differently. Grammatical factors are implemented as grammatical constraints within HG. As such, their model maintains the basic properties of HG. Specifically, the constraints place substantive linguistic limits on possible variation, making it a grammar dominant model. Unlike earlier generative models, and similar to the Labovian approach, grammar also contributes to the frequency with which different variants are observed. On the other hand, since non-grammatical factors can only move the weights of existing constraints up or down but not add or remove constraints, these factors can influence the frequency with which different forms are observed, but cannot add to or remove from the set of grammatical forms.

Coetzee and Kawahara (2013) applied their model successfully to one non-grammatical factor (usage frequency) and to two variable phenomena (English *t/d*-deletion and Japanese geminate devoicing). It is unclear, however, how generalizable the model is to both other grammatical phenomena and other non-grammatical factors. The current paper addresses this by applying their model to non-grammatical factors other than usage frequency, and to a different variable process. Specifically, this paper investigates the interaction of phonological grammar with speech rate and with perceptual priming/inhibition on the perception of variable cross-word nasal place assimilation in English (i.e. *green boat* realized as *gree[n] boat* or *gree[m] boat*).

The rest of this paper is structured as follows: Section 2 reviews the existing literature on cross-word nasal place assimilation in English, focusing on the perceptual compensation for such assimilation. Section 3 discusses two speech perception experiments conducted to investigate the interaction of phonological grammar, speech rate and perceptual inhibition/priming in the perception of place assimilated nasals, and section 4 develops an account of the results in the Coetzee/Kawahara noisy HG model. Finally, section 5 concludes.

## 2. Cross-word nasal place assimilation in English

Since the focus of the paper is on the proposed model of phonological variation rather than on a specific variable process, a process that is fairly well understood was selected for further investigation, namely the perception of cross-word nasal place assimilation in English. For the same reason two non-grammatical factors that are known to influence speech processing were selected, namely speech rate and perceptual inhibition by priming. Nasal place assimilation is one of the most common phonological processes in the world's languages. In English, it applies obligatorily word-internally (*a[m]ber*, *\*a[n]ber*) and variably across word boundaries (*gree[n] box* ~ *gree[m] box*). The rest of this paper will focus only on cross-word assimilation.

### 2.1. Production patterns of cross-word nasal place assimilation

Most of the research on the production of cross-word nasal place assimilation has focused on whether assimilation is complete or incomplete, and evidence has been found for both possibilities. In terms of acoustic measures, for instance, Gow (2001, 2002, 2003) has found evidence that the assimilation can be incomplete, while Zimmerer *et al.* (2009) and Dilley and Pitt (2007) report evidence of acoustically complete assimilation. Similarly, there is also articulatory evidence for both incomplete (e.g. Barry 1985; Kerswill 1985; Ellis and Hardcastle 2002) and complete (Ellis and Hardcastle 2002) assimilation.

In languages that has non-optional nasal place assimilation processes, /n/ assimilates either in both pre-labial and pre-velar context, or only pre-velar context—no language has nasal place assimilation in pre-labial context only (2002:194, 375-381; Mohanan 1993:76-77). The source of the velar assimilation preference is most likely to be found in perception. Various discrimination studies have found that [n] and [ŋ] are more confusable than [n] and [m] (Narayan 2006:53; Wang and Bilger 1973:1254; etc.). If place assimilation of /n/ originates in the misperception of [n] (Ohala 1990, 1993), it can be expected that /n/ will assimilate more in contexts where the result will be more confusable with [n]. This same pre-velar preference is also observed in variable nasal place assimilation, with higher assimilation rates observed in pre-velar than pre-labial contexts. For German, Zimmerer *et al.* (Zimmerer *et al.* 2009:Table II & III), for instance, report pre-velar assimilation rates of 21.5% and pre-labial assimilation rates of only 13.0%. Dilley and Pitt (2007) investigate the assimilation of word-final /n/ in the *Buckeye Corpus* (Pitt *et al.* 2007). Although they do not report assimilation separately for pre-labial and pre-velar contexts, Dilley reports in personal communication that pre-velar contexts had higher assimilation rates (21.8% vs. 18.5%) (Dilley 2013).

In addition to being influenced by grammatical factors such as place of assimilation, nasal place assimilation is also influenced by non-grammatical factors such as speech rate. Kerswill (1985) and Barry (1985, 1991) both investigated cross-word place assimilation in English using electropalatography, and found higher rates of assimilation at faster than slower speech rates. Kerswill, for instance, reports for one speaker an assimilation rate of 50% when speaking at a normal rate, and 90% when speaking at a fast rate (1985:31, Table 2).

Under the assumption that perception will reflect the patterns observed in production, two hypotheses can be formulated about how listeners will compensate for nasal place assimilation. Based on the assumed independence between grammatical and non-grammatical factors (see section 1), a third hypothesis can be added. All three of these hypotheses are confirmed in the experiments presented in section 3.

#### *Hypothesis 1: a place effect*

Given that assimilation is more likely in pre-velar than pre-labial contexts, listeners will be more likely to compensate for assimilation in pre-velar contexts.

*Hypotheses 2: a rate effect*

Given that assimilation is more likely at faster than slower speech rates, listeners will be more likely to compensate for assimilation at faster speech rates.

*Hypothesis 3: independence of place and rate effects*

The effect of speech rate on the perception of assimilated nasals will be the same in velar and labial contexts.

## 2.2. Perception of cross-word place assimilated nasals

Most of the research on the perception of nasal place assimilation has focused not on how listeners compensate for place-assimilated nasals, but rather on the likelihood that listeners would misperceive alveolar nasals as labial/velar when followed by a labial/velar consonant (i.e. that listeners would perceive an utterance like [...nb...] as /...mb.../) (Beddor and Evans-Romaine 1995; Hura *et al.* 1992; Ohala 1990; etc.). Though relevant to understand how processes of place assimilation originates, this research is less relevant to the experiments discussed in section 3, since these experiments present listeners with assimilated nasals and ask how listeners compensate for this assimilation.

There is a body of literature that shows that listeners do compensate for nasal place assimilation during perception. These results are based on experimental paradigms as diverse as word identification (e.g. asking listeners whether a phrase like [grim bɑks] contains the word *green*; Darcy *et al.* 2009), to phoneme identification (asking listeners to identify a consonant ambiguous between [m] and [n] in a phrase like *lea[m/n] bacon*; Gaskel and Marslen-Wilson 1998), to priming (asking whether an assimilated nasal word primes its unassimilated citation form—i.e. does *lea[m]* in *lea[m] bacon* prime *lean*; Gaskel and Marslen-Wilson 1996, 2001; Gow 2002; Gow and McMurray 2007). Although the details differ between these various studies, all of them find evidence that listeners do compensate for appropriately produced nasal place assimilation—that is listeners are more likely to identify a token like *te[m]* as an utterance of *ten* in a pre-labial (*te[m] pens*) than a pre-velar (*te[m] cars*) or pre-alveolar (*te[m] toes*) context. Given results such as these, we can expect to find evidence of compensation for assimilation in the experiments reported below.

Hypothesis 2 above stated that more compensation for assimilation is expected at faster than slower speech rates. There is also evidence that listeners do adjust their perceptual strategies based on speech rate. Most of this research focuses on the differentiation between segments that contrast in the temporal domain. Nooteboom (1981:144), for instance, shows that the decision boundary between short and long /a/ in Dutch depends on speech rate. Similar results have been reported for the distinction between short lag and long lag VOT plosives in English (Port 1977; Port and Dalby 1982), the distinction between geminate and singleton consonants in Japanese (Sonu *et al.* 2013), etc.

More directly relevant to the current paper is research showing listeners to be sensitive to the influence of speech rate on the application of phonological processes. Function words in English are subject to variable realization under different speech rates. In slow speech, the *or* in a phrase like *leisure or time* is usually realized, while it can be completely absent at faster rates. Dilley and Pitt (2010) show that listeners are sensitive to this correlation between speech rate and the realization of function words. Specifically, listeners are more likely to perceive the function word *or* in faster than slower speech rates. An small eye-tracking study by Li and Kaiser (2012) report similar results for cross-word nasal place assimilation in English. They first taught the participants in their study to associate the made-up names *Vone~Vome*, *Kine~Kime* and *Shoon~Shoom* with the images of six imaginary creatures. They then presented their participants with sentences containing one of these six names, with the task to look at the image corresponding to the name. In the crucial sentences, these names were followed by labial-initial words—i.e. contexts in which the alveolar nasals in *Vone*, *Kine* and *Shoon* could assimilate to the following labial. For instance, the names were presented in sentences like “*Every time the waiter brings out a strawberry*

*cheesecake from the kitchen, Vome peeks to see if he can steal a piece.*" They varied the speech rate of the context (the non-underlined part of the sentence), and found that listeners were significantly more likely to look at the image corresponding to the alveolar-final names (*Vone, Kine, Shoon*) at faster than slower speech rates. This shows that their listeners were more likely to assume the speaker intended a final alveolar nasal, but assimilated it in place to the following labial at faster speech rates. Results such as these show that listeners are sensitive to the correlation between speech rate and the application of phonological processes, adding additional support to Hypothesis 2.

The Li/ Kaiser and Dilley/Pitt studies manipulated speech rate of a whole frame sentence, giving listeners a large sample of speech over which to calculate speech rate. What remains unknown is whether listeners require this much evidence to compensate for speech rate. Would their results have been the same if they had not embedded the crucial parts of their stimuli in a frame sentence, but rather presented just the crucial part at different speech rates? In the experiments reported in section 3 below, the tokens were presented in isolation (not embedded in larger sentence frames) so that this question can be answered.

There are several ways in which listeners' performance can be influenced by the task they are asked to perform. Gerrits and Schouten (2004; also Pitt and Samuel 1993), for instance, review how the results a phoneme identification experiments can differ based solely on the format of the task that listeners have to perform (two alternative forced choice, ABX discrimination, same-different, etc.). Different kinds of tasks place more or less demands on short term memory, taps into pre-lexical or lexical processing, prime or inhibit competitors, etc., and all of these impact how listeners perform. In the experiments reported below, the same stimuli will be presented in two different task conditions. Both experiments will present listeners with /n/-final words followed by words that start on plosives at different places of articulation. The /n/-final words will be produced with place assimilation to the following plosives (e.g. *te[m] pens*). In Experiment 1, listeners will be asked whether the stimulus contained a production of the relevant English word (i.e. *ten*). In Experiment 2, listeners will be asked to choose from among the three possible pronunciations of the /n/-final word (*ten~tem~teng*). In Experiment 1, a high percentage of 'yes' responses is expected. Listeners are not explicitly made aware of the possibility of the assimilated pronunciations and would therefore not go into the task considering these pronunciations as hypotheses. In Experiment 2, however, the alternatives are presented explicitly as possible answers, and listeners will therefore approach the task with these possibilities as active hypotheses. A lower percentage of alveolar (*ten*) responses is expected in Experiment 2 than the percentage 'yes' responses in Experiment 1. By making listeners aware of the assimilated pronunciations, these percepts are primed as hypotheses, which are therefore expected to inhibit the unassimilated, alveolar percept. This leads to a fourth hypothesis. The assumed independence of grammatical and non-grammatical effects gives the fifth hypothesis.

*Hypothesis 4: an inhibition effect*

Listeners will be more likely to compensate for assimilation when only the unassimilated token is presented as an option than when unassimilated and assimilated tokens are presented as options.

*Hypothesis 5: independence of the inhibition and place effects*

The effect of inhibition on the perception of assimilated alveolar nasals will be the same in velar and labial contexts.

A note is in order about how the experiments relate to listeners perceptual experiences in real life outside of the laboratory. Listeners are likely to be confronted with speech at different rates so that it is easy to see why the hypothesis with regard to speech rate is relevant to everyday communicative contexts. However, listeners are rarely placed in a context such as that in Experiment 2 where they have to choose between different pronunciations of the same word. There are, however, instances in which the communicative circumstances require listeners to pay careful attention to the fine acoustic details of an

utterance—this is what Lindblom *et al.* (1995) call “how-mode” or “signal oriented” perception as opposed to the more usual “what-mode” or “content oriented” perception. Listeners may resort to the “how-mode” in adverse listening conditions, when trying to identify the accent of a speaker, when interacting with a second language speaker, etc. Though the default is the “what-mode”, there are contexts in which fine phonetic details will be attended to and weighted more heavily. Additionally, even if this kind of listening were only rarely encountered in normal communicative situations, if listeners perform differently in Experiment 2 from Experiment 1 in the laboratory, it would still be evidence about the cognitive system that underlies the perception of variable acoustic inputs. A complete model of listeners’ perceptual competence would still have to account for such differences.

### 3. Experiments

This section of the paper discusses two experiments on the perception of cross-word nasal place assimilation in English, designed to test the five hypotheses stated above. Both experiments test Hypotheses 1 and 2 (on the place effect and the rate effect) and Hypothesis 3 (on the independence of the place and rate effects). The comparison between the two experiments tests Hypothesis 4 (on the inhibition effect) and Hypothesis 5 (on the independence of the place and inhibition effects).

#### 3.1. Experiment 1: *n*?

##### 3.1.1 Methods

###### *Token selection*

Ten /n/-final English words were selected, and each of them was combined with a word starting on an alveolar, labial and velar plosive to form a two-word. Our main interest is in the comparison of the labial and velar contexts, and the phrases in these contexts were therefore selected to be matched for collocational likelihood. The likelihood of the word *ten* given that the next word is *cat* (or *cats* or *cat’s*) was calculated in the *Corpus of Contemporary American English* (Davies 2008-). The likelihood of *ten* given that the next word was *pen* (or *pen* or *pen’s*) were similarly calculated. Similar values were calculated for all labial/velar pairs. A pair-wise *t*-test was performed on these values, finding no evidence for a difference between the contexts ( $t(9) = 0.50$ , two-tailed  $p = 0.63$ ). In addition to these tokens, five filler words that end on consonants other than /n/ were also selected, and used in three two-word phrases, with no specific limitation on the sound on which the second word could start. The full list of two-word phrases is given in the table in (1).

(1) Token stimuli

Word	Pre-Alveolar	Pre-Velar	Pre-Labial
aspirin	aspiri[n] tablet	aspiri[ŋ] capsule	aspiri[m] powder
ten	te[n] toes	te[ŋ] cats	te[m] pens
Ben	Be[n] Thomas	Be[ŋ] Kingsley	Be[m] Potter
pen	pe[n] tip	pe[ŋ] case	pe[m] pocket
van	va[n] tires	va[ŋ] keys	va[m] price
bargain	bargai[n] deal	bargai[ŋ] getaways	bargai[m] books
billion	billio[n] dollars	billio[ŋ] gallons	billio[m] barrels
cannon	canno[n] defense	canno[ŋ] guards	canno[m] balls
John	Joh[n] Doe	Joh[ŋ] Green	Joh[m] Black
muffin	muffi[n] top	muffi[ŋ] cups	muffi[m] pans

#### Filler stimuli

Word	Correct pronunciation	Incorrect pronunciation	Contexts
bitter	[bɪtɪ]	[bɪrl]	__apple, __coffee, __orange
copper	[kɑpɪ]	[kɑpl]	__bullet, __ore, __crown
dog	[dɑg]	[dɑp]	__house, __food, __fur
face	[feɪs]	[feɪf]	__plunge, __cream, __time
fast	[fæst]	[fæsk]	__answer, __run, __reply

#### *Stimulus creation*

Multiple repetitions of each of the token and filler phrases were recorded by a phonetically trained, male native speaker of American English. Recordings were made with *Praat* (Boersma and Weenink 2013) in a sound attenuated room, at a sampling rate of 44.1 kHz. Phrases were all produced with appropriate assimilation in the pre-velar and pre-labial contexts—i.e. *pen case* was produced as *pe[n] case*. Filler tokens were recorded with the correct final consonant, and with one incorrectly produced (but acoustically related) final consonant—see the bottom table in (1).

To identify /n/-final tokens that were produced with clear [n], [ŋ] and [m], a token selection experiment was conducted. For each /n/-final word, those three repetitions that were judged to have the most clear [n], [ŋ] and [m] were identified. Each of these words were then spliced out of their two-word phrases and presented in isolation four times in random order to eight native speakers of American English. The participants' task was to indicate whether a token ended on [n], [m] or [ŋ]. That repetition of each token that was most often identified accurately (i.e. in agreement with actual acoustic properties of the token) was selected for use in the main experiment. The average percent correct identification for the selected tokens was 98%.

One repetition of each of the two pronunciations of the filler words was also selected, and spliced out of its two-word phrase. A repetition of the second word in each two-word phrase, both for the token and filler words, was selected. Each first word was then combined with its appropriate second word. In order to make the assimilated [ŋ]- and [m]-tokens sound natural, no silence was included between the first and second words when they were spliced together. Since the experimental task focuses on the initial word, the intensity of the initial word in each two-word phrase was scaled to an average of 75 dB using *Praat's* "Scale intensity ..." command, while the second word was scaled to 70 dB.

Two additional copies of each recombined phrase were resynthesized using the Pitch-Synchronous Overlap and Add (PSOLA) algorithm in *Praat*. One copy was resynthesized to be 80%, and the other 60% of the duration of the original phrase. The PSOLA algorithm keeps in tact the spectral properties of a token, altering only the speech rate.

The tables in (1) list the filler and token stimuli used in the experiment. For each two-word phrase, three copies were included in the experiment (slow, faster, and fastest rate) giving a total of 90 token phrases, and 90 filler phrases.

#### *Participants*

Twenty-three undergraduate students from the University of Michigan were recruited as participants. All participants were native speakers of American English who self-reported having no speech or hearing deficits. Participants were paid for their participation.

#### *Procedure*

Token presentation and response collection were controlled with *Superlab 4* stimulus presentation software ([www.superlab.com](http://www.superlab.com)). Auditory tokens were presented over headphones at a comfortable listening level. Responses were collected through a response box connected to a MacBook laptop

computer. The experiment was conducted in a sound attenuated room. The 180 two-word phrases were each presented five times, differently randomized for each participant. Simultaneous with the auditory presentation of a phrase, participants were presented on the monitor with the first word of the corresponding two-word phrase, represented with its normal, accepted spelling—i.e. even when the auditory token did not contain an accurate pronunciation of the first word, the visual stimulus was the word in its actual spelling. The task of participants was to indicate whether the phrase contained an utterance of the word presented by pushing a button on the response box. Participants were instructed to respond fast without sacrificing accuracy.

### Hypotheses

The hypotheses were motivated in section 2 above. *Hypothesis 1 (place)* stated that an effect of place of articulation is expected. More “yes” responses in velar than labial contexts would be support for this hypothesis. *Hypotheses 2 (rate)* predicted that compensation should be more likely at faster speech rates. More “yes” responses at the faster speech rates would provide support for this hypothesis. *Hypothesis 3 (independence of place of rate)* was that the effects of rate and place should be independent. According to this hypothesis the effect of speech rate should therefore be equal in labial and velar contexts.

### 3.1.2 Results

Figure 1 summarizes the results of Experiment 1. Participants were very accurate at identifying the tokens that end on alveolar consonants – accuracy was 99%, 99% and 98% at the slow, faster and fastest rates. Since the main focus is on how listeners perceived place assimilated nasals, alveolar tokens were excluded from further analysis. Figure 1 also shows that participants were more likely to perceive an alveolar final token in velar than labial contexts (90% vs. 83%). Lastly, the figure shows a rate effect, with more alveolar percepts at faster rates (slow = 80%, faster = 87%, fastest = 93%).

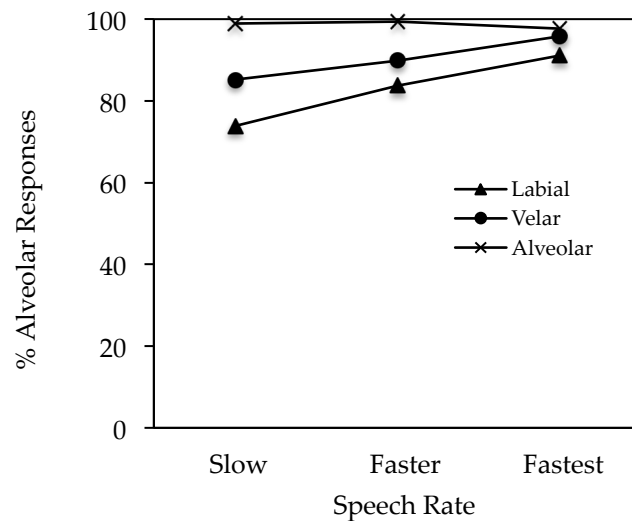


Figure 1: Response patterns in Experiment 1 ( $n?$ ). The  $y$ -axis represents the percent “yes” responses, and the  $x$ -axis speech rate. Places of articulation is indicated with differently marked lines.

In order to assess the influence of place of articulation and speech rate on the responses a linear mixed effects model was fit to the data, using the `lmer()` function from the R `lme4` package (Bates *et al.* 2013). Since the response variable (yes/no) was binary, a binomial link function was used. Place of articulation (POA: velar, labial), speech rate (Rate: slow, faster, fastest) and their interaction were entered as factors, with random intercepts for participant and word, and random slopes for participants by POA and Rate. The interaction between POA and Rate was not found to be significant, and a simpler model

excluding the interaction term was run. The results of this model are summarized in the table in (2). In this model, “labial” is the reference value for POA, and “slow” for Rate. The results for the intercept are hence those for the labial context at the slow rate, with all of the other results expressing the relation to this category. The `lmer()` function does not test for the main effects of the two factors, but rather for the individual differences between the reference level and the other levels of each factor. However, since POA has only two levels, the significant difference found between the reference level “labial” and “velar” shows that there is a main effect for this factor.<sup>1</sup> Since Rate has three levels, the results of the `lmer()` function cannot be used to determine whether there is a main effect of Rate. For this reason, a second model was fit to the data, excluding Rate. The results of the two models (with and without Rate) were then compared using the `anova()` function in R. This comparison shows that the model with Rate fits the data significantly better ( $\chi^2(9) = 656.79$ ,  $p < 0.001$ ), confirming that there is also a main effect of Rate.

(2) Results of linear mixed effects model for Experiment 1

	Estimate	Standard Error	z-value	p
Intercept	0.208	0.422	0.492	0.610
POA(velar)	0.672	0.238	2.821	< 0.005
Rate(faster)	0.701	0.083	8.416	< 0.001
Rate(fastest)	1.590	0.300	12.251	< 0.001

Experiment 1 therefore provides support for *Hypotheses 1 (place)* in finding more compensation in pre-velar than pre-labial contexts, and for *Hypothesis 2 (rate)* in finding more compensation at faster rates. Though other research (reviewed in section 2.2) has shown that listeners can adjust their perceptual strategies to variation in speech rate, this earlier research has used longer utterances from which listeners could calculate rate. That perceptual adjustment based on rate is found in this experiment even for phrases consisting of only two words, shows that listeners need very little evidence to calculate rate and also that they can make virtually immediate perceptual adjustments to changes in speech rate. Given the lack of evidence for a significant interaction between place and rate, *Hypothesis 3 (independence of rate and place)* is also supported.

### 3.2. Experiment 2: *n/m/ng?*

#### 3.2.1 Methods

##### *Token selection*

Five /n/-final words were selected as tokens for this experiment. As in Experiment 1, each word was used in two-word phrases, with the second word starting on an alveolar, labial or velar plosive. Additionally, twelve other words were selected as filler items. Fillers all ended on a sound other than /n/, and was also

---

<sup>1</sup> A question that should be considered is whether this place effect truly has its origin in a difference between the two places of articulation or whether it may be epiphenomenal of some kind of frequency effect. The fact that listeners were more likely to identify *ten* in the phrase *te[n] cats* than in the phrase *te[n] pens* may be because *ten cats* is a more common phrase than *ten pens*. As shown in section 3.1, however, collocational probabilities of the /n/-final words with the following word were equal in the velar and labial conditions. Another possibility is that the biphone probability of the vowel in the initial word and [m] is higher than that of the vowel and [ŋ]. If this were the case, then listeners may have had more /m/ percepts (hence less compensation for assimilation) than /ŋ/ percepts as a result of this frequency-based bias. Evaluating this possibility is difficult, and with currently available corpora, impossible. Corpora that are used to calculate biphone probabilities are based on citation form pronunciations. These corpora would include a word like *ten* in calculating the biphone probability of the sequence [ɛn], but not in calculating the probabilities of sequences like [ɛŋ] or [ɛm]. Given how often /n/ assimilates in place to a following consonant, these probabilities would therefore not give an accurate measure of the actually experienced biphone probabilities of any of the sequences [ɛn], [ɛŋ], or [ɛm]. This would be problematic especially in evaluating the results of experiments that focus specifically on the production of words like *ten* in contexts where assimilation is likely.

used as the first word in a two-word phrase. There was no specific limit on the sound in which the second word in the filler phrases started. The full set of tokens is given in (3).

**Stimulus creation**

Stimulus creation was done as in Experiment 1. Since the two-word phrases with /n/-final words used in Experiment 2 were also used in Experiment 1, the same recordings of these words were reused in Experiment 2. What was different from Experiment 1, is that the [n], [m] and [ŋ]-final pronunciations of the /n/-final words was spliced into each of the three contexts—i.e. all three pronunciations were used in a two-word phrase before a labial, alveolar and velar plosive. All of these phrases that were created are listed in the top table in (3). The twelve filler phrases were recorded multiple times by the same speaker, using the same procedures. The first word in each filler phrase was produced with both the correct pronunciation, and with the final consonant replaced with a different but acoustically related consonant—see the bottom table in (3). One production of each filler word was selected that was judged to be a clear production of the intended pronunciation. A clear production of each second word was also selected. As with the rest of the stimuli, the first words were scaled to an intensity of 75 dB and the second to 70 dB.

There was a total of 45 phrases containing /n/-final words: 5 words, each produced with a final [n], [m] and [ŋ], in pre-alveolar, pre-velar, and pre-labial contexts. There were 24 additional filler phrases—12 words, with two pronunciations of each, used with the same second word. As in Experiment 1, three copies were created of each of these 69 phrases: at the original (slow) rate, and at a faster (80% of the original) and an even faster rate (60% of the original).

(3) /n/-final tokens

Word	Pre-Alveolar	Pre-Velar	Pre-Labial
aspirin	aspiri[n]/[m]/[ŋ] tablet	aspiri[n]/[m]/[ŋ] capsule	aspiri[n]/[m]/[ŋ] powder
ten	te[n]/[m]/[ŋ] toes	te[n]/[m]/[ŋ] cats	te[n]/[m]/[ŋ] pens
Ben	Be[n]/[m]/[ŋ] Thomas	Be[n]/[m]/[ŋ] Kingsley	Be[n]/[m]/[ŋ] Potter
pen	pe[n]/[m]/[ŋ] tip	pe[n]/[m]/[ŋ] case	pe[n]/[m]/[ŋ] pocket
van	va[n]/[m]/[ŋ] tires	va[n]/[m]/[ŋ] keys	va[n]/[m]/[ŋ] price

Other tokens

Word	Correct pronunciation	Incorrect pronunciation
ball	[bal] game	[baʊ] game
bitter	[bɪtɪ] pill	[bɪl] pill
dress	[dɹɛs] code	[dɹɛf] code
face	[feɪs] mask	[feɪf] mask
honor	[hɑnɪ] code	[hɑnl] code
knife	[naɪf] blade	[naɪs] blade
meal	[mi:l] time	[mi:ɪ] time
press	[pɹɛs] release	[pɹɛf] release
rough	[ɹɑf] time	[ɹɑs] time
safe	[seɪf] house	[seɪs] house
silver	[sɪlvɪ] dollar	[sɪvl] dollar
tool	[tu:l] kit	[tur] kit

**Participants**

Eighteen undergraduate students from the University of Michigan were recruited as participants for the experiment. All participants were native speakers of American English, who self-reported having no speech or hearing deficits. Participants were paid for their participation in the experiment.

### **Procedure**

Stimulus presentation and response collection were done as in Experiment 1. Tokens were presented in the same random order for all participants. Phrases with /n/-final words where the place of articulation of the /n/ agreed with the onset of the following word were presented 10 times (*te[n] toes*, *te[m] pens*, *te[ŋ] cats*). The labial and velar tokens were the focus of the experiment, since this is where participants are expected to accommodate for place assimilation. All other tokens functioned as fillers. Phrases with a mismatch in place of articulation between the nasal and onset of the following word were each presented twice. The remaining phrases were each presented three times. The total number of tokens was therefore 756.

The task in Experiment 2 was different from that in Experiment 1. For phrases that contain an /n/-final word, participants were presented with three choices, indicating the three pronunciations of the /n/-final word. For example, simultaneous with the presentation of the auditory token *te[m] pens*, participants would see ‘ten/teng/tem’ on the monitor. Their task was to indicate which of these pronunciations was used in the auditory stimulus by pushing a button on the response box. The first visual token was always the correct pronunciation ending on [n], the second the [ŋ]-pronunciation, and the third the [m]-pronunciation. For the other phrases, the visual presentation contained only two possible options, corresponding to the two possible pronunciations of the first word in these phrases.

### **Hypotheses**

This section briefly explains how the hypotheses motivated in section 2 are tested in this experiment, and what response patterns are expected under each of the hypotheses. *Hypothesis 1 (place)* would be supported if participants select the [n]-final answer more often in the velar than the labial contexts. In accordance with *Hypothesis 2 (rate)*, more [n]-final answers are expected at faster speech rates. *Hypothesis 3 (independence of place and rate)* would receive support if the effect of speech rate is found to be equal between labial and velar contexts. *Hypothesis 4 (inhibition)* and *Hypothesis 5 (independence of inhibition and place)* are tested by comparison of the results between the two experiments—see section 3.3.

### **3.2.2 Results**

Figure 2 visualizes the results of Experiment 2. This figure shows the percentage of responses where listeners selected the alveolar-final option. Only phrases where the final nasal of the first word agrees in place with the onset of the second word are included—i.e. *te[n] toes*, *te[m] pens*, and *te[ŋ] cats*. As in Experiment 1, participants were very accurate at identifying alveolar-final tokens in the alveolar contexts (97%, 96% and 90% at the slow, faster and fastest rates). These tokens were excluded from further analysis. The results for the labial and velar contexts are very similar as in Experiment 1, except that overall alveolar responses were lower in Experiment 2. In agreement with *Hypothesis 1 (place)*, participants were more likely to perceive an alveolar final token in velar than labial contexts (40% vs. 28%). There was also an effect of rate as predicted by *Hypothesis 2 (rate)*, with more alveolar responses at faster rates (slow = 20%, faster = 32%, fastest = 50%).

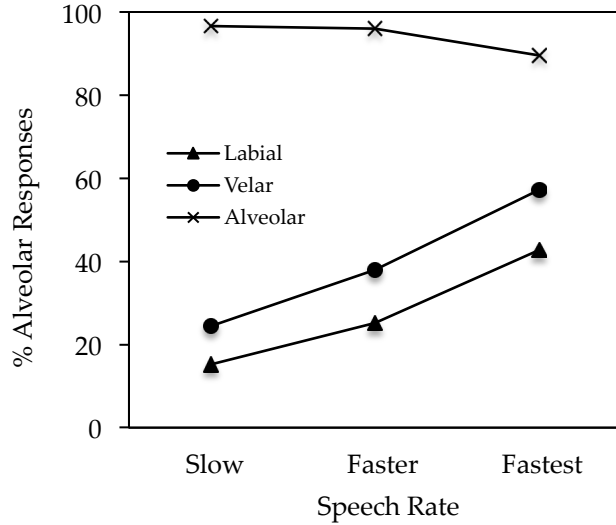


Figure 2: Response patterns in Experiment 2 (*n/m/ng?*). The *y*-axis represents the percent alveolar-final responses, and the *x*-axis speech rate. Places of articulation are indicated with differently marked lines.

The same model used in the analysis of Experiment 1 was fit to the results of Experiment 2. As before, the interaction between Rate and POA was not found to be significant, so that the final model did not include this term. The results of this model are summarized in the table in (4), with “labial” is the reference value for POA, and “slow” for Rate. As in Experiment 1, these results show a main effect for POA. To assess the contribution of Rate, a second model was fit, excluding Rate as factor, and the results of the two models were compared using the `anova()` function in R. This comparison confirms that the model with Rate fits the data significantly better than the model without Rate ( $\chi^2(9) = 460.79$ ,  $p < 0.001$ ).

(4) Results of linear mixed effects model for Experiment 2

	Estimate	Standard Error	z-value	p
Intercept	-2.090	0.327	-6.395	< 0.001
POA(velar)	0.778	0.282	2.761	< 0.006
Rate(faster)	0.711	0.099	7.160	< 0.001
Rate(fastest)	1.667	0.131	12.708	< 0.001

Like Experiment 1, Experiment 2 provides support for both *Hypotheses 1 (place)* in finding more compensation in pre-velar than pre-labial contexts, and for *Hypothesis 2 (rate)* in finding more compensation at faster speech rates. Given the lack of a significant interaction between place and rate, *Hypothesis 3 (independence of rate and place)* is also supported. *Hypothesis 4 (inhibition)* and *Hypothesis 5 (independence of inhibition and place)* are evaluated next by comparing the results between the experiments.

### 3.3. Comparing Experiments 1 and 2

Visual inspection of Figure 1 and Figure 2 shows that the response patterns in the two experiments were similar—more alveolar responses in velar than labial contexts, and at faster speech rates in both experiments. There is, however, one way in which the results of the experiments diverge. Participants in Experiment 1 (*n?*) were much more likely to have an alveolar percept than those in Experiment 2 (*n/m/ng?*) (87% vs. 34%), in accordance with *Hypothesis 4 (inhibition)*. The figures also show that both velar and labial contexts were impacted similarly by the inhibition of alveolar percepts in Experiment 2 (*n/m/ng?*), suggesting that the inhibition effect is independent from the place effect, in accordance with *Hypothesis 5 (independence of inhibition and place)*.

To investigate these patterns in more detail, a linear mixed effects model was fit to the data. The model contained the same factors as those used in the individual experiments, plus an additional two-level factor for Experiment (Exp: Exp1, Exp2). As before, the model contained random intercepts for word and participant, and random slopes for participants (by POA, Rate and Exp). A model with interactions between Exp, Rate and POA was initially fit to the data, but none of the interaction terms were found to contribute significantly. A simpler model without interaction terms was therefore fit. The reference level for POA was “labial”, for Rate it was “slow”, and for Exp it was “Exp2”. The results of this model are summarized in the table in (5).

(5) Linear mixed effects model for the combined results of Experiments 1 and 2

	Estimate	Standard Error	z-value	p
Intercept	-1.795	0.321	- 5.605	< 0.001
POA(velar)	0.683	0.238	2.875	< 0.005
Rate(faster)	0.715	0.083	8.668	< 0.001
Rate(fastest)	1.648	0.128	12.900	< 0.001
Exp(Exp2)	3.549	0.3337	10.551	< 0.001

The significant effect of the Exp factor in the model provides confirmation for *Hypothesis 4 (inhibition)*. When the /m/-final and /ŋ/-final percepts are primed, the /n/-final percept is inhibited, resulting in significantly fewer alveolar-final percepts. The lack of an interaction between the inhibition and place effects provides support for *Hypothesis 5 (independence of inhibition and place)*, adding additional evidence that grammatical and non-grammatical factors tend to be independent from each other. Lastly, the lack of evidence for an interaction between the rate and inhibition effects shows that these two non-grammatical factors are also independent from each other.

#### 4. A noisy Harmonic Grammar account

This section of the paper develops a computational model of the listener based on the results collected in the two experiments discussed above. The model will be developed in noisy Harmonic Grammar (noisy HG; Coetzee and Pater 2011; Jesney 2007), and more specifically in the version of noisy HG developed by Coetzee and Kawahara (2013) that allows the weights of faithfulness constraints to be scaled by non-grammatical factors. This scaled version of noisy HG allows for formal interaction of grammatical and non-grammatical factors in one model. The contribution of grammatical factors is captured by the use of grammatical constraints, and the contribution of non-grammatical factors by the scaling of faithfulness constraint weights. Specifically, a non-grammatical factor that promotes application of a process scales down the weight of faithfulness constraints. This results in faithfulness constraints contributing less to the harmony score of unfaithful candidates so that these candidates will have higher harmony scores, and therefore a higher likelihood of being selected as output. Factors that inhibit application of a process will do the opposite.

##### 4.1. The data

Two possible pronunciations will be considered for each token word, the faithful pronunciation with an [n], and a pronunciation with an appropriately assimilated nasal (i.e. [m] in a labial context and [ŋ] in a velar context). In Experiment 1 (*n?*), it is not possible to know how listeners identified the word if they responded “no”. If a listener was presented with a token like *te[m] pens*, and responded “no”, this listener could have identified the word as *te[m]* or *te[ŋ]*. It is highly unlikely, though, that a listener would have identified the word as *te[ŋ]*. First, such a percept would have no acoustic support from the stimulus. Secondly, *te[ŋ] pens* would be an ungrammatical pronunciation of *ten pens*. In Experiment 2 (*n/m/ng?*), when presented with *te[m] pens*, participants could choose between “ten/teng/tem”, so that we can

determined how often listeners selected the token with an inappropriate place of articulation. When presented with a labial token, participants selected the velar response just over 1% of the time, and when presented with a velar token, they selected the labial response only 3.4% of the time. Since these responses represent less than 2% of all the data, they are not included in the model developed below.

In developing the initial model, I abstract away from the influence of speech rate and lexical inhibition. This model is therefore based on the percent alveolar responses in the velar and labial contexts, as averaged across the different rates and the two experiments. The table in (6) shows the average frequency across all three rates and both experiments, at which listeners selected the faithful and assimilated percepts.

(6) Percentage of different responses, averaged across the two experiments and three speech rates

	Faithful percept		Assimilated percept	
Velar context	'te[n] cats'	35	'te[ŋ] cats'	65
Labial context	'te[n] pens'	45	'te[m] pens'	55

#### 4.2. The basic grammar

Harmonic Grammar (HG; Pater 2009; Smolensky 2006), like Optimality Theory (OT), is a constraint-based model of grammar. The HG account of the experimental results relies on the constraints defined in (7). The markedness constraint AGREEPLACE is the driving force behind the place assimilation. MAX[F]/DEP[F] rather than IDENT[F] constraints will be used to regulate faithfulness. Since the assimilation rates differ in labial and velar contexts, the constraints must be able to differentiate between these contexts, so that IDENT[place] would not be sufficient. McCarthy and Prince already hinted at this approach to featural faithfulness in 1994 (McCarthy and Prince 1994: footnote 9), while Zoll (1996) represents the first formal development of this approach (albeit in a pre-Correspondence Theory PARSE/FILL form).

- (7)
- AGREEPLACE Assign one violation mark for every nasal consonant that differs in place from a directly following consonant.
  - MAX[alveolar] Assign one violation mark for every [alveolar] feature in the Input that does not have a correspondent in the Output.
  - DEP[labial] Assign one violation mark for every [labial] feature in the Output that does not have a correspondent in the Input.
  - DEP[velar] Assign one violation mark for every [velar] or [labial] feature in the Output that does not have a correspondent in the Input.

There is a cross-linguistic implicational relationship between labial and velar triggers of nasal place assimilation: There are languages without nasal place assimilation, languages in which only velars trigger assimilation, and languages in which both velars and labials trigger assimilation. However, there are no languages in which only labials trigger assimilation. A possible perceptual origin of this implication was presented in section 2.1. Such universal implicational generalizations have been accounted for in two different ways in constraint-based phonology. Prince and Smolensky suggested very early on that some constraints may be in a universally fixed ranking (Prince and Smolensky 1993; published in 2004). To capture the implicational place assimilation DEP[labial] could be stipulated to universally outrank DEP[velar]. However, stipulating universally fixed ranking (or weighting) relationships between constraints undermines some of the core assumptions of constraint-based phonology where every ranking (or weighting) of constraints is usually assumed to be a possible human language. De Lacy (2002, 2004, 2006) suggest an alternative approach, showing that such generalizations can be captured by defining constraints as being in a stringency relationship. This approach of De Lacy's is adopted here. A

candidate with an acquired [labial] feature therefore violates both DEP[labial] and DEP[velar]. On the other hand, a candidate with an acquired [velar] feature violates only DEP[velar].

Unlike in OT where constraints are ranked, constraints are weighted in HG. A harmony score (H) is then calculated for each candidate, based on the constraints that it violates and the weights of those constraints, according to the formula in (8). The candidate with the highest H-score is selected as the output. The specific formula used here is that of noisy HG (Coetzee and Pater 2011; Jesney 2007) rather than classic HG (Pater 2009). Noisy HG is to classic HG as stochastic OT (Boersma and Hayes 2001) is to classic OT. The only difference between classic HG and noisy HG, is that the weight of each constraint is perturbed by a small normally distributed, random amount of noise every time the grammar is used. Because of these random perturbations the H-scores for a given candidate will be slightly different every time that the grammar is used, potentially resulting in variation.

$$(8) \quad H(cand) = \sum_{i=1}^n (w_i + nz_i) C_i(cand)$$

Where  $C_i$  is the  $i$ th constraint,  $w_i$  is the weight of this constraint,  $nz_i$  the noise associated with  $C_i$  at a given evaluation, and  $C_i(cand)$  the number of  $C_i$ -violations of candidate  $cand$ , expressed as a negative integer.

In order to determine the weights that would be needed for the constraints in (7) to produce the frequency patterns in (6), a learning simulation was run using a noisy HG learning algorithm implemented in *Praat* (Boersma and Weenink 2013). For details on the learning algorithm see Boersma and Pater (2008) and Coetzee and Pater (2008). The input file for the learning simulation assumed that each of the labial and velar contexts was encountered 100 times, with the place assimilation rates given in (6).<sup>2</sup> In running the simulation, the “decision strategy” was set to “PositiveHG”.<sup>3</sup> All other settings were kept at *Praat*’s defaults. Due to the noisiness of evaluation, slightly different weights will be learned each time that the algorithm is used. The simulation was therefore run ten times, and the average weights across the ten simulations were used in the rest of the paper. These weights are given in the top line of the tableau in (9),<sup>4</sup> which also shows the violations earned by each candidate. The final column gives the H-score of each candidate, not taking into account the noisiness of the evaluation.

(9) Constraint weights and H-scores

		133.82	66.18	66.18	-27.43	
		AGRPLACE	MAX[alv]	DEP[vel]	DEP[lab]	H-score
Velar	/...n k.../    [...n k...]	-1				-133.82
	[...ŋ k ...]		-1	-1		-132.36
Labial	/...n p.../    [...n p...]	-1				-133.82
	[...m k ...]		-1	-1	-1	-133.36

As can be seen in this tableau, the H-scores of the assimilation and faithful candidates are very close in both the velar and labial contexts. Consequently, once noise is factored into the evaluation, the relation between the H-scores of the two candidates may flip from one evaluation occasion to the next, resulting in variation. Secondly, in both the velar and labial contexts, the H-score of the assimilation candidate is higher than that of the faithful candidate, so that assimilation should be more likely. Comparison with the

<sup>2</sup> The learning input file is available from the author upon request.

<sup>3</sup> Several different versions of the HG learning algorithm have been implemented in *Praat*. The algorithm called “PositiveHG” in *Praat* is used in this paper. In this algorithm, any constraint weight lower than 1 is replaced by 1 when harmony scores are calculated. The simulations reported in this paper were also run using the “LinearOT” algorithm, where negative constraint weights are replaced by zero. There was no appreciable difference in performance of the models learned with the two different algorithms.

<sup>4</sup> See the previous footnote on the various ways in which negative constraint weights are handled in Harmonic Grammar.

frequencies in (6) shows that this agrees with the frequency patterns in the learning data. Lastly, the H-score of the assimilation candidate in the velar context is higher than that in the labial context, implying that assimilation should be observed more frequently in velar than labial contexts—again in agreement with the patterns in (6). It can therefore be expected that this grammar will mimic the relative frequency patterns in (6). In order to verify this, *Praat*'s "To output distribution ..." function was used (with all settings at their default values) to determine the output patterns that this grammar would generate. Due to the noisiness of the evaluation, slightly different output patterns are predicted each time that the function is used. For this reason, the procedure was repeated 10 times, and the average output patterns over the 10 repetitions were calculated. The rate of assimilation predicted by this grammar is compared to the actually observed rates in (10), showing that the grammar accounts well for the observed frequency patterns.

(10) Observed and expected rates of assimilation (in percentage)

		Faithful percept		Assimilated percept	
		O	E	O	E
Velar context	'te[n] cats'	35	33.7	'te[ŋ] cats'	65 66.3
Labial context	'te[n] pens'	45	44.8	'te[m] pens'	55 55.2

Figure 3 plots the performance of the participants in the two experiments and the predictions of the HG model. Although the HG model captures the grammatical component of the response patterns (the difference between labial and velar contexts), it does not capture the effect of either speech rate or lexical inhibition. The next section introduces weight scaling as a method of incorporating the contribution of these two factors into the model.

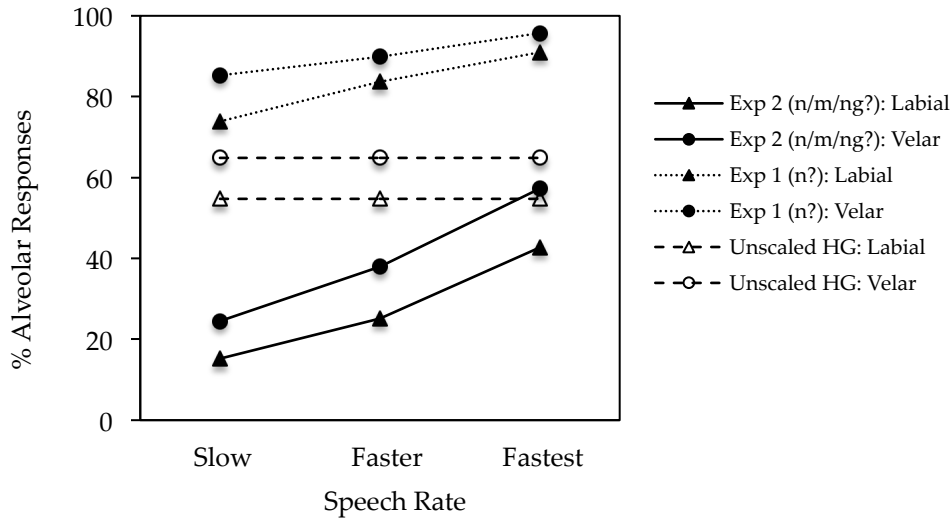


Figure 3: Comparing the performance of the unscaled HG model to the performance of the participants. Circles mark labial and triangles mark velar responses. Filled markers indicate performance of the participants, and open markers the predictions of the HG model. Solid lines mark performance of participants in Experiment 2 (*n/m/ng?*), and the stippled lines in Experiment 1 (*n?*).

### 4.3. Scaling the grammar

The rate of nasal assimilation predicted by the model developed in section 4.2 is higher than the actual rate in Experiment 2 ( $n/m/ng?$ ) and lower than the actual rate in Experiment 1 ( $n?$ ). In Experiment 2 ( $n/m/ng?$ ), being unfaithful is less likely than predicted by the HG model, which means that the faithfulness constraints should contribute more towards determining the H-scores in modeling the results of this experiment. Following Coetzee and Kawahara (2013), this will be implemented by scaling the weights of faithfulness constraints up from those learned in section 4.2. Conversely, the weight of faithfulness constraints will be scaled down when modeling the results of Experiment 1 ( $n?$ ). The effect of rate will be accommodated in the same manner, by scaling faithfulness down for faster rates and up for slower rates.

#### 4.3.1 Modeling inhibition

The table in (11) shows the assimilation rates, averaged across different speech rates, for each of the experiments, as well as averaged between the two experiments. The amount with which the faithfulness constraints should be scaled up or down for the two experiments was determined by an iterative best-fit procedure. For Experiment 1 ( $n?$ ) where faithfulness constraints need to be scaled down, the basic grammar weights learned above in section 4.2 were scaled down in increments of 0.05. The output pattern predicted by each of these scaled versions of the grammar was then determined using Praat's "To output distribution ..." function in exactly the same manner as described above in section 4.2. The mean square error (MSE)<sup>5</sup> of these predicted assimilation rates relative to the rates actually observed in Experiment 1 ( $n?$ ) was then calculated. This procedure was repeated until the scaling factor was found that resulted in the smallest MSE. This scaling factor was found to be -1.45 for Experiment 1 ( $n?$ ). A similar procedure was used for Experiment 2 ( $n/m/ng?$ ), except that the basic weights from section 4.2 were scaled up in increments of 0.05. The scaling factor that resulted in a minimization of the MSE for this experiment was found to be 1.20. The tables in 0 show the predicted assimilation rates for a few different scaling factors for the two experiments, together with the MSE's associated with these scaling factors. Each table shows the rates for the scaling factor that minimizes the MSE for the particular experiment, as well as the scaling factor just below and just above this value. These tables show the effect of weight scaling: as the weights of the faithfulness constraints are scaled down, being unfaithful (i.e. assimilating) becomes more likely, with the opposite happening as the weights are scaled up. The updated H-score formula, incorporating this scaling factor, is given in (13). In this formula,  $sf_{inhibition}$  stands for the inhibition scaling factor, and is replaced by -1.45 for Experiment 1 ( $n?$ ) and 1.20 for the Experiment 2 ( $n/m/ng?$ ).

- (11) Comparison between the overall rates of assimilation (in percentage), and the assimilation rates in Experiment 1 ( $n?$ ) and Experiment 2 ( $n/m/ng?$ ).

	Average	Experiment 1 ( $n?$ )	Experiment 2 ( $n/m/ng?$ )
Velar context	65	90	40
Labial context	55	83	28

<sup>5</sup> MSE is calculated by determining the difference between the predicted and observed assimilation rate for the velar and labial contexts, respectively, and then squaring each of these differences. The MSE is the average between the squared errors for the velar and labial contexts.

- (12) The influence of different scaling factors on the predicted assimilation rates. The scaling factor is added to the weights of the faithfulness constraints learned in section 4.2 before the output patterns of the scaled grammars are calculated.

a. Experiment 1 ( $n?$ )

Scaling Factor	Observed		Predicted		MSE
	Velar	Labial	Velar	Labial	
-1.40	90	83	89.04	82.61	0.54
-1.45	90	83	89.55	83.42	0.19
-1.50	90	83	90.07	84.09	0.60

b. Experiment 2 ( $n/m/ng?$ )

Scaling Factor	Observed		Predicted		MSE
	Velar	Labial	Velar	Labial	
1.15	40	28	40.43	29.81	1.73
1.20	40	28	39.34	28.82	0.55
1.25	40	28	38.19	27.74	1.67

$$(13) \quad H(cand) = \sum_{i=1}^n (w_i + nz_i + sf_{inhibition}) F_i + \sum_{j=1}^m (w_j + nz_j) M_j$$

Where  $F_i$  is the  $i$ th faithfulness constraint,  $w_i$  is the weight of this constraint,  $nz_i$  the noise associated with this constraint at this particular evaluation occasion,  $sf_{inhibition}$  the inhibition scaling factor, and  $F_i(cand)$  the number of times that candidate  $cand$  violates  $F_i$ , expressed as a negative integer. And where  $M_j$  is the  $j$ th markedness constraint,  $w_j$  is the weight of this constraint,  $nz_j$  the noise associated with the constraint at this particular evaluation occasion, and  $M_j(cand)$  the number of times that candidate  $cand$  violates  $M_j$ , expressed as a negative integer.

The predictions of this scaled model are much closer to the observed rates of assimilation in each of the experiments than the unscaled model from section 4.2. In fact, in terms of MSE, the scaled model represents an improvement of 88% over the unscaled model (unscaled MSE = 793.42, scaled MSE = 95.37).<sup>6</sup> However, this scaled model does not take into account the effect of speech rate, and therefore predicts the same rates of assimilation for all three speech rates. In the next section, this model is augmented to also incorporate the contribution of speech rate to assimilation.

### 4.3.2 Speech rate

The effect of speech rate is incorporated in the same way as lexical inhibition. Faithfulness constraints are scaled down for faster speech rates, resulting in higher rates of unfaithfulness (assimilation), and conversely for slower speech rates. In determining the scaling factors to be used for speech rate, the augmented model developed in section 4.3.1 was used as starting point. For Experiment 1 ( $n?$ ), the starting point is therefore the weights of the faithfulness constraints scaled down by -1.45 from the basic values learned in section 4.2. Similarly, for Experiment 2 ( $n/m/ng?$ ), the starting values are the basic values from section 4.2 scaled up by 1.20.

Different scaling factors in increments of 0.05 were tested for each of the three speech rates (slow, faster, fastest). The output patterns predicted by each of the grammars based on the 0.05 incremental scaling factors were determined by submitting each grammar ten times to Praat's "To output distributions ..." function. The average assimilation rates for velar and labial contexts across these ten runs were taken to be the assimilation rates predicted by that particular scaling factor. As before, the MSE

<sup>6</sup> The MSE's reported here are calculated using the actual assimilation rates, i.e. without averaging across different speech rates. The percentage improvement is calculated as follows:  $\frac{MSE(unscaled\ model) - MSE(scaled\ model)}{MSE(unscaled\ model)} \times 100$ .

was calculated for each of the incremental 0.05 scaling factors. After doing this separately for each of the two experimental conditions, the MSE’s associated with a specific scaling factor in the two experiments were averaged. For each speech rate, the scaling factor associated with the smallest MSE averaged between the two experiments was taken to be the scaling factor associated with that particular speech rate.

- (14) Assimilation rates predicted at different rate scaling factors for the “faster” speech rate in each of the two inhibition conditions. The column corresponding to the rate scaling factor that minimizes the overall MSE is shaded. This marks the scaling factor that will be used in the noisy HG model developed here.

		Rate Scaling Factors						Actual
		-0.10	-0.05	0.00	0.05	0.10	0.15	
Exp 1 (n?)	Labial	84.70	84.10	83.42	82.63	81.89	81.11	84
	Velar	90.53	90.06	89.55	89.07	88.54	87.91	90
	MSE	0.66	0.07	0.12	0.98	2.65	5.46	
Exp 2 (n/m/ng?)	Labial	30.76	29.71	28.82	27.73	26.86	25.87	25
	Velar	41.45	40.37	39.34	38.20	37.04	36.02	38
	MSE	21.89	13.36	7.75	3.46	1.99	2.25	
Overall MSE		11.28	6.72	3.94	2.22	2.32	3.85	

The table in (14) illustrates this procedure. This table shows the assimilation rates predicted at the “faster” rate at different rate scaling factors for each of the experimental conditions. The MSE is also given for each scaling factor and each experiment separately. Finally, the last row in the table contains the mean MSE for the two experiments at a particular scaling factor. The column that corresponds to the scaling factor that results in the lowest mean MSE for the two experiments is shaded in grey. For the “faster” rate, this scaling factor (0.05) will be used as scaling factor for the “faster” rate in the model developed here. The scaling factor that minimizes the MSE for each of the other two speech rates was determined in the same manner. This factor was found to be -0.70 for the “fastest” rate, and 0.65 for the “slow” rate.

### 4.3.3 Summary

The formula in (15) gives the final model, including the scaling factors for both inhibition and speech rate. The grammar that was learned is summarized in the tables in 0. The constraint weights and scaling factors from this table can be substituted into the formula in (15) to find the form of the grammar to be used for any combination of speech rate and inhibition condition. Figure 4 compares the performance of this final scaled model against the response patterns of the participants in the experiment. Comparison of this figure with Figure 3 shows that the scaled model fits the observed data much better. The fully scaled model, in fact, represents an improvement of over 99% over the unscaled model.<sup>7</sup>

$$(15) \quad H(cand) = \sum_{i=1}^n (w_i + nz_i + sf_{inhibition} + sf_{rate}) F_i(cand) + \sum_{j=1}^m (w_j + nz_j) M_j(cand)$$

Where  $F_i$  is the  $i$ th faithfulness constraint,  $w_i$  is the weight of this constraint,  $nz_i$  the noise associated with the constraint at this particular evaluation occasion,  $sf_{inhibition}$  the inhibition scaling factor,  $sf_{rate}$  the rate scaling factor, and  $F_i(cand)$  the number of times that candidate  $cand$  violates  $F_i$ , expressed as a negative integer. And where  $M_j$  is the  $j$ th markedness constraint,  $w_j$  is the weight of this constraint,  $nz_j$  the noise associated with the constraint at this particular evaluation occasion, and  $M_j(cand)$  the number of times that candidate  $cand$  violates  $M_j$ , expressed as a negative integer.

<sup>7</sup> See footnote 6 for how the improvement is calculated.

(16) Summary information of the model

a. Constraint weights

Constraint	Weight
AGREEPLACE	133.82
MAX[alveolar]	66.18
DEP[velar]	66.18
DEP[labial]	-27.38

b. Scaling factors

Priming/Inhibition	Scaling factor
Experiment 1 ( <i>n?</i> )	-1.45
Experiment 2 ( <i>n/m/ng?</i> )	1.20

Speech rate	Scaling factor
Slow	0.65
Faster	0.05
Fastest	-0.70

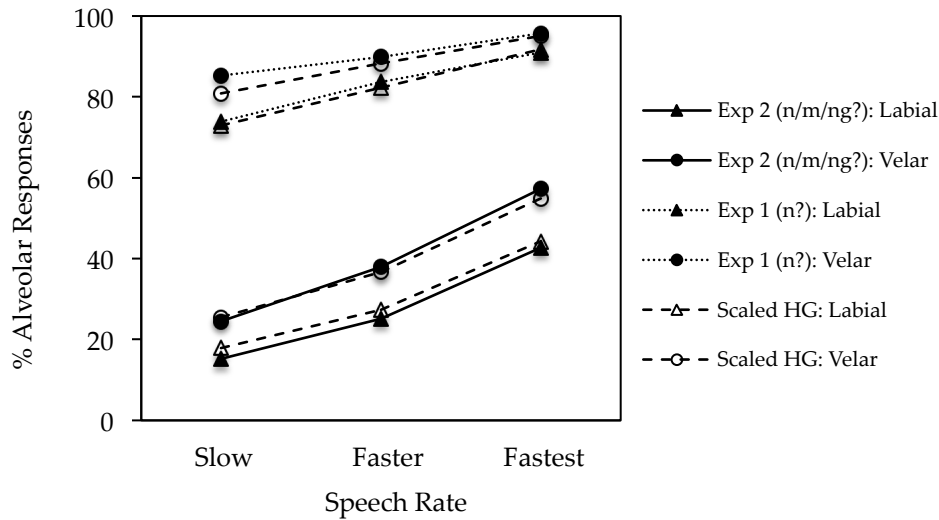


Figure 4: Comparing the performance of the fully scaled HG model against the participant response patterns. Circles mark labial responses, and triangles mark velar responses. Filled markers indicate performance of the participants, and open markers the predictions of the scaled HG model. The solid lines marks performance of participants in the Experiment 2 (*n/m/ng?*), and the stippled line performance in Experiment 1 (*n?*).

5. Concluding discussion

Given that phonological performance is determined by a complex collaboration between grammar and several non-grammatical factors, any sufficient model of the language user’s phonological competence must have a mechanism for integrating all of the different factors that contribute towards that competence. Following in the footsteps of Coetzee and Kawahara (2013), this paper presented such a model within noisy Harmonic Grammar. This section discusses some of properties of this model, considers ways in which the model could be extended, and also points out some shortcomings of the model.

## 5.1. Grammar dominance

As discussed in section 1, in the generative tradition it has been assumed that, though both grammatical and non-grammatical factors co-determine the properties of actually produced speech, grammar takes precedence over non-grammatical factors. Grammar defines the space of possible or grammatical forms, and the non-grammatical forms can only affect how frequently the forms that grammar allows are observed. The non-grammatical factors cannot override grammar or introduce new forms that are deemed ungrammatical.

Within the domain of phonological variation, there is another way in which grammar can be dominant. In addition to encoding what the possible grammatical forms are, grammar can also impose a specific frequency relationship between two variable processes. Consider for instance variable *t/d*-deletion from word-final clusters in English. This process has been documented for many different varieties of English, and as evidenced in the reviews of relevant research in Labov (1989) and Coetzee (2004; also Coetzee and Pater 2011; Coetzee and Kawahara 2013), though dialects may differ in the absolute rate of deletion in different phonological contexts, in all dialects deletion before a consonant (*west bank*) is more likely than before a vowel (*west end*). A grammar dominant model can do more than just specify that deletion is possible before both vowels and consonants, but can also specify that deletion before a consonant will always be more likely than before a vowel. In a constraint-based grammar (whether that be OT or HG) these frequency limitations follow from the specific constraints that are used to model the variable phenomenon. In modeling variable *t/d*-deletion, for instance, Coetzee and Kawahara (2013) assume a general anti-deletion constraint MAX that is violated by deletion in any context, and a position specific constraint that is violated only by deletion in pre-vocalic position MAX-PRE-V. Since deletion in pre-consonantal context violates a proper sub-set of the constraints violated by deletion in pre-vocalic position (only MAX vs. MAX and MAX-PRE-V), it follows that deletion in pre-consonantal context will imply deletion in pre-vocalic context, but not the other way around. Anttila and Andrus (2006) show that this type of implicational generalization (that they call a “T-Order”) holds true of variation too such that the possible deletion rates in pre-consonantal context is guaranteed to be always at least as high as that in pre-vocalic context (i.e. no language with more deletion in pre-vocalic position is possible). Coetzee and Pater (2011: section 4.4) confirmed this prediction by showing that a constraint-based grammar assuming MAX and MAX-PRE-V fails to successfully learn a grammar that can produce more deletion pre-vocalically than pre-consonantly, even when presented with learning data that has more deletion pre-vocalically.

In the account of nasal place assimilation developed here, grammar similarly specifies a specific frequency relationship between assimilation in a pre-labial and a pre-velar context. The faithfulness constraints used in the account developed here were defined such that assimilation in a velar context will always violate a proper subset of the constraints violated by assimilation in a labial context—see (7). As shown in (17), the consequence is that the assimilation candidate in velar contexts will always have a higher harmony score than the assimilation candidate in labial contexts. Neither the specific weights of the constraints, nor any way in which the constraint weights may be scaled by non-grammatical factors can change this relationship between the two assimilation contexts.

(17) Harmony scores of assimilation in pre-velar and pre-labial contexts

	MAX [alv]	DEP [vel]	DEP [lab]	H
/...n k.../ → [...ŋ k...]	-1	-1		$-(w\text{MAX}[\text{alv}] + w\text{DEP}[\text{vel}])$
/...n p.../ → [...m k...]	-1	-1	-1	$-(w\text{MAX}[\text{alv}] + w\text{DEP}[\text{vel}] + w\text{DEP}[\text{lab}])$

Since grammar dominance depends on both the architecture of the grammar (how constraints interact) and the specific constraints selected, it follows that not all constraint-based grammars will have this property. For instance, if DEP[velar] were defined such that it was violated only by acquisition of a velar feature and not also by acquisition of a labial feature, a language with more assimilation in pre-labial than pre-velar context would be possible. Grammar dominance therefore depends partially on the meta-theory about constraints, and specifically on the strong assumption that constraints are universal and responsible for determining linguistic typology. It is the typological observation that assimilation in pre-labial position implies assimilation in pre-velar context that motivated the specific stringency formulation of the constraints DEP[labial] and DEP[velar]. From the way that these constraints are defined follows the grammar dominance illustrated here. Grammar dominance hence follows from assuming a strong Universal Grammar, and from a grammatical model in which Universal Grammar places substantive limits on possible languages.

In another sense, grammar dominance depends specifically on the architecture of the model assumed here. In this model, grammatical and non-grammatical factors are treated very differently. Grammatical factors are encoded into the model via specific grammatically motivated constraints, and as shown just above, it is these constraints that are responsible for setting the limits of what a possible language is. Non-grammatical factors are implemented as merely scaling the weights associated with the grammatical constraints. Non-grammatical factors are not constraints, and neither can they add or remove constraints. If non-grammatical factors were implemented as constraints on par with grammatical constraints, or if they were allowed to remove constraints, then the strong typological limits placed on possible languages by the set of grammatical constraints will no longer hold true.

Grammar dominance is not a property of models of phonological variation that do not assume a strong Universal Grammar. As shown in section 1 above, the variable rule approach as implemented by Cedergren and Sankoff (1974) and widely used in variationist sociolinguistics does not differentiate between grammatical and non-grammatical factors. The same is true of some implementations of usage-based models (Bybee 2001, 2006; 2007; etc.), or exemplar models (Gahl and Yu 2006 and papers therein; Pierrehumbert 2001; etc.). Most of these models make no formal distinction between grammatical and non-grammatical factors. In fact, in describing usage-based grammar, Bybee first defines the usage-based conceptualization of grammar as “the cognitive organization of one’s experience with language” (Bybee 2006:711). Later on the same page she describes how this organization is done as follows: “... the general cognitive capabilities of the human brain, which allow it to categorize and sort for identity, similarity, and difference, go to work on the language events a person encounters, categorizing and entering in memory these experiences.” Grammar is the result of cognitive organization achieved with general cognitive abilities, not with grammar or language specific abilities. Exactly the same cognitive abilities that organize our experience with social interactions and with our physical environment organize our experience with language. No formal distinction is made between how language and other aspects of our experience are processed or stored in the mind. If a child acquiring a language were to be exposed to a set of experiences where assimilation happens to be observed more often in pre-labial than pre-velar context, the general abilities of the mind to classify would notice this pattern, and codify this as the grammar. This view of grammar is fundamentally different from the type of approach advocated in this paper. Under the approach pursued here, there are language specific cognitive capacities (Universal Grammar represented in the constraint set, as well as in the principles for how constraints interact via their weights). Language is processed according to these principles and not with general cognitive capabilities. This places a limit on the types of grammars that can be learned.

There is relatively little research on how grammatical and non-grammatical factors co-determine the patterns of phonological variation. The data currently available point mostly toward independence between grammatical and non-grammatical factors, and to grammar dominance, and the model

presented here has therefore been developed to have these properties. However, more research is needed that specifically probes into the relationship between grammatical and non-grammatical factors in phonological variation. If independence or grammar dominance is found not to hold true upon more research, the model presented here will have to be adapted accordingly.

## 5.2. More on scaling factors

In the model developed above certain assumptions were made about how scaling factors are implemented that may be too simplistic. First, the two scaling factors (for the rate and inhibition effects) were implemented as additive. Given that no interaction was found between the rate and inhibition effects in the two experiments (see section 3.3) it was warranted to implement these two effects as independent and additive. However, it is likely that various non-grammatical factors may interact with each other in complex manners. In the variationist sociolinguistic tradition it has in fact been established that biographical characteristics of speakers can interact such that age or socioeconomic status may have a different effect for men than women. It has also been found in this tradition that biographical characteristics of speakers may interact with things like speech style or register (see for instance the discussion on *r*-deletion in section 1). A fully adequate model of phonological variation will therefore have to include a meta-theory about how various non-grammatical factors interact with each other. One option, for instance, may be to have a separate model that combines the influence of all non-grammatical factors, and then to have only a single scaling factor in the model that reflects the combined contribution of all non-grammatical factors. More research is needed, however, about which non-grammatical factors do interact, and also about the nature of the interactions that are observed. Only once more is known about these issues, will it be possible to develop a more sophisticated model of the contribution of non-grammatical factors to phonological variation.

A second assumption is that scaling factors affect faithfulness constraints only, and that they scale all faithfulness constraints alike (the weights of all faithfulness constraints are moved up or down by the same quantity)—this is in agreement with the proposal of Van Oostendorp (1997; see also Itô and Mester 2001). This is most likely a too simplistic approach. It is conceivable, for instance, that application of one variable process and non-application of another process may be associated with a specific socio-economic status. In such case, members of a speech community who belong the specific socio-economic status group would have grammars where some faithfulness constraints are scaled down and others up. Similarly, it is not the case that all variable phenomena will be affected in the same way by things like speech rate. Though it is likely that all variable articulatory simplification phenomena (deletion, assimilation, consonantal lenition, etc.) are more likely to apply at faster than slower speech rates, the opposite pattern may well be observed in variable augmentation processes such as variable epenthesis (see Auger 2001 on Vimeu Picard; Nevins 2007 on Brazilian Portuguese). See Coetzee and Kawahara (2013: section 5.2) for similar remarks with regard to usage frequency and variable simplification vs. augmentation processes.

The model proposed (but not developed) by Boersma and Hayes (2001: appendix C) is more suited to handling a single non-grammatical factor having different impacts on different constraints. Their proposal is implemented in stochastic OT rather than noisy HG, but the basic principle is the same. They assume that non-grammatical factors (they focus on style/register, but their proposal could easily be extended to other factors as well) can move constraints up or down the hierarchy, but crucially they assume that this is decided on a constraint-by-constraint and register-by-register basis. For each speech register, it has to be determined for each constraint whether or not it has to be moved at all, and if it has to be moved whether it has to be moved up or down the hierarchy and by how much. This is obviously a more powerful model of how non-grammatical factors can contribute to phonological variation than the one developed in this paper.

The correct model is most likely somewhere between that developed here and the Boersma/Hayes-proposal. What is needed is a meta-theory about which constraint families are affected how by different non-grammatical factors. For instance, if it is found that processes that add structure (e.g. epenthesis) are affected in one way, and processes that remove structure (e.g. deletion, assimilation) are affected in a different way by speech rate or usage frequency, then this difference has to be built into the model. More research is needed into how non-grammatical and grammatical factors co-determine patterns of phonological variation. Only once we understand these issues better, will it be possible to develop a well-motivated model of how non-grammatical factors should be incorporated into a model of phonological variation. Until such time, however, the most restrictive model consistent with the data at hand is to be preferred.

### 5.3. Final remark

The treatment of variation in phonological theory has come a long way over the past 50 years. Where initially it was barely acknowledged and treated as something mostly in the domain of performance and hence not part of the concern of linguistics, variation is now a central part of phonological theory. Through decades of research in variationist sociolinguistics, the speech perception/production tradition and theoretical phonology, much is known about how phonological variation works and how it can be modeled. The time is now ripe for synthesizing, for developing a more complete model that brings together all the factors that co-determine phonological variation. This paper is one proposal for what such a comprehensive model of phonological variation could look like. This model has been applied with success to the contribution of usage frequency to phonological variation by Coetzee and Kawahara (2013), and to the contribution of speech rate and perceptual inhibition to variation in the current paper. The model developed here is not the final word on the topic. There are many ways in which this model falls short. As more research is conducted into how grammatical and non-grammatical factors codetermine the patterns of phonological variation, the model presented here needs to be augmented, adapted, or possibly even abandoned. It is my hope that this paper will be just one conversational turn in an ongoing conversation about how to account for phonological variation in a cognitively realistic manner.

### References

- Anttila, Arto. (1997). Deriving variation from grammar. In Hinskens, Frans, Roeland van Hout & Leo Wetzels (eds.) *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 35-68.
- Anttila, Arto. (2002). Variation and phonological theory. In Chambers, Jack K., Peter Trudgill & Natalie Schilling-Estes (eds.) *Handbook of Language Variation and Change*. Oxford: Blackwell, 206-243.
- Anttila, Arto. (2006). Variation and opacity. *Natural Language and Linguistic Theory* 24, 893-944.
- Anttila, Arto & Curtis Andrus. (2006). *T-Orders*. Ms. Stanford University.
- Anttila, Arto, Vivienne Fong, Stefan Benus & Jennifer Nycz. (2008). Variation and opacity in Singapore English consonant clusters. *Phonology* 25, 181-216.
- Auger, Julie. (2001). Phonological variation and Optimality Theory: evidence from word-initial vowel epenthesis in Vimeu Picard. *Language Variation and Change* 13, 253-303.
- Bane, Max. (2011). Deriving the structure of variation from the structure of non-variation in the English dative alternation. In Washburn, Mary Byram, Katherine McKinney-Bock, Erika Varis, Ann Sawyer & Barbara Tomaszewicz (eds.) *Proceedings of the 28th annual meeting of the West Coast Conference on Formal Linguistics*. Somerville: Cascadilla Press, 42-50.
- Barry, Martin C. (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics* 4, 1-16.
- Barry, Martin C. (1991). Temporal modeling of gestures in articulatory assimilation. *Proceedings of the XIIth International Congress of Phonetic Sciences*. Aix-en-Provence: University of Provence, 14-17.

- Bates, Douglas, Martin Maechler & Ben Bolker. (2013). *lme4: Linear mixed-effects models using Eigen and Eigen++, R package version 0.999999-2*. <http://CRAN.R-project.org/package=lme4> (Last viewed June 9, 2013).
- Bayley, Robert. (2002). The quantitative paradigm. In Chambers, J.K, Peter Trudgill & Natalie Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Oxford: Blackwell, 117-141.
- Beddor, Patrice Speeter & David Evans-Romaine. (1995). Acoustic-perceptual factors in phonological assimilations: a study of syllable-final nasals. *Rivista di Linguistica* 7, 145-174.
- Boersma, Paul. (1997). How we learn variation, optionality, and probability. *University of Amsterdam Institute of Phonetic Sciences Proceedings* 21, 43-58.
- Boersma, Paul & Bruce Hayes. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.
- Boersma, Paul & Joe Pater. (2008). *Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar*. Ms. University of Amsterdam and University of Massachusetts, Amherst. <http://www.fon.hum.uva.nl/paul/papers/boersmaPaterHGGLA.pdf>. Accessed 11 July 2012.
- Boersma, Paul & David Weenink. (2013). *Praat: Doing Phonetics by Computer (Version 5.3.56)*. [Computer Program.] Retrieved May 23, 2012, from <http://www.praat.org>.
- Bybee, Joan L. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, Joan L. (2006). From usage to grammar: the mind's response to repetition. *Language* 82, 711-733.
- Bybee, Joan L. (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.
- Cedergren, Henrietta J. & David Sankoff. (1974). Variable rules: Performance as a statistical reflection of competence. *Language* 50, 333-355.
- Chomsky, Noam. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam & Morris Halle. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- Coetzee, Andries W. (2004). *What It Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. Ph.D., Dissertation, University of Massachusetts.
- Coetzee, Andries W. (2006). Variation as accessing "non-optimal" candidates. *Phonology* 23, 337-385.
- Coetzee, Andries W. (2009a). An integrated grammatical/non-grammatical model of phonological variation. In Kang, Young-Se, Jong-Yurl Yoon, Hyunkyung Yo, Sze-Wing Tang, Yong-Soon Kang, Youngjun Jang, Chul Kim, Kyoung-Ae Kim & Hye-Kyung Kang (eds.) *Current Issues in Linguistic Interfaces. Volume 2*. Seoul: Hankookmunhwa, 267-294.
- Coetzee, Andries W. (2009b). Phonological variation and lexical frequency. In Schardl, Anisa, Martin Walkow & Muhammad Abdurrahman (eds.) *NELS 38. Volume 1*. Amherst: GLSA, 189-202.
- Coetzee, Andries W. (2012). Variation: where laboratory and theoretical phonology meet. In Cohn, Abigail C., Cécile Fougeron & Marie Huffman (eds.) *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, 62-75.
- Coetzee, Andries W. & Shigeto Kawahara. (2013). Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31, 47-89.
- Coetzee, Andries W. & Joe Pater. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26, 289-337.
- Coetzee, Andries W. & Joe Pater. (2011). The place of variation in phonological theory. In Goldsmith, John, Jason Riggle & Alan Yu (eds.) *Handbook of Phonological Theory. 2nd Edition*. Cambridge: Blackwell, 401-434.
- Darcy, Isabelle, Franck Ramus, Anne Christophe, Katherine Kinzler & Emmanuel Dupoux. (2009). Phonological knowledge in compensation for native and non-native assimilation. In Kügler, Frank, Caroline Féry & Ruben van de Vijver (eds.) *Variation and Gradience in Phonetics and Phonology*. Berlin: Walter de Gruyter, 265-310.
- Davies, Mark. (2008-). *The Corpus of Contemporary American English: 450 Million Words, 1990-Present*. Available online at <http://corpus.byu.edu/coca/>.
- de Lacy, Paul (2002). *The Formal Expression of Markedness*. Ph.D., University of Massachusetts.

- de Lacy, Paul. (2004). Markedness and conflation in Optimality Theory. *Phonology* 21, 145-199.
- de Lacy, Paul. (2006). *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.
- Dilley, Laura. (2013). *E-mail message received on October 30, 2013*.
- Dilley, Laura C. & Mark A. Pitt. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America* 122, 2340–2353.
- Dilley, Laura C. & Mark A. Pitt. (2010). Altering context speech rate can cause words to appear and disappear. *Psychological Science* 21, 1664-1670.
- Ellis, Lucy & William J. Hardcastle. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30, 373-396.
- Fasold, Ralph W. (1991). The quiet demise of variable rules. *American Speech* 66, 3-21.
- Gahl, Susanne & Alan Yu. (eds.) (2006). *Linguistic Review. Volume 23(3). Special Issue on Exemplar-Based Models in Linguistics*. Berlin: De Gruyter Mouton.
- Gaskel, M. Gareth & William D. Marslen-Wilson. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22, 144-158.
- Gaskel, M. Gareth & William D. Marslen-Wilson. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24, 380-396.
- Gaskel, M. Gareth & William D. Marslen-Wilson. (2001). Lexical ambiguity resolution and spoken word recognition: bridging the gap. *Journal of Memory and Language* 44, 325-349.
- Gerrits, Ellen & M.E.H. Schouten. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics* 66, 363-376.
- Gow, David W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45, 133-159.
- Gow, David W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28, 163-179.
- Gow, David W. (2003). Feature parsing: feature cue mapping in spoken word recognition. *Perception & Psychophysics* 65, 575-590.
- Gow, David W. & Bob McMurray. (2007). Word recognition and phonology: The case of English coronal place assimilation. In Cole, Jennifer S. & Jose I. Hualde (eds.) *Papers in Laboratory Phonology 9*. New York: Mouton de Gruyter, 173-200.
- Guy, Gregory R. (1991). Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* 3, 1-22.
- Hayes, Bruce & Colin Wilson. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379-440.
- Hura, Susan L., Björn Lindblom & Randy L. Diehl. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35, 59-72.
- Itô, Junko & Armin Mester. (2001). Covert generalizations in Optimality Theory: the role of stratal faithfulness constraints. *Proceedings of the 2001 International Conference on Phonology and Morphology*. Seoul: The Phonology-Morphology Circle of Korea, 3-33.
- Jesney, Karen. (2007). *The locus of variation in weighted constraint grammars*. Ms. Poster presented at the Workshop on Variation, Gradience and Frequency in Phonology. Stanford University, July 2007. [Downloaded on December 27, 2007 from <http://people.umass.edu/kjesney/papers.html.%5D>].
- Kay, Paul & Chad K. McDaniel. (1979). On the logic of variable rules. *Language in Society* 8, 151-187.
- Kerswill, Paul E. (1985). A sociophonetic study of connected speech processes in Cambridge English: An outline and some results. *Cambridge Papers in Phonetics and Experimental Linguistics* 4.
- Labov, William. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

- Labov, William. (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45, 715-762.
- Labov, William. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. (1989). The child as linguistic historian. *Language Variation and Change* 1, 85-97.
- Li, David Cheng-Huan & Elsi Kaiser. (2012). Accounting for Speech Rate in Spoken Word Recognition. *InterSpeech 2012: Proceedings of the 13th Annual Conference of the International Speech Communication Association*. 2007-2010.
- Lindblom, Bjorn, Susan Guion, Susan L. Hura, Seung-Jae Moon & Raquel Willerman. (1995). Is sound change adaptive? *Rivista di Linguistica* 7, 5-37.
- McCarthy, John J. & Alan Prince. (1994). The emergence of the unmarked: optimality in prosodic morphology. In González, Mercè (ed.) *Proceedings of NELS 24*. Amherst: GLSA, 333-379.
- Mendoza-Denton, Norma, Jennifer Hay & Stefanie Jannedy. (2003). Probabilistic sociolinguistics: beyond variable rules. In Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.) *Probabilistic Linguistics*. Cambridge, MA: MIT Press, 97-138.
- Mohanan, Karuvannur P. (1993). Fields of attraction in phonology. In Goldsmith, John (ed.) *The Last Phonological Rule: Reflections on Constraints and Derivations*. Chicago: Chicago University Press, 61-116.
- Narayan, Chandan R. (2006). *Acoustic-perceptual salience and developmental speech perception*. Ph.D., Dissertation, University of Michigan.
- Nevins, Andrew. (2007). Review of Sheer (2004). *Lingua* 118, 425-434.
- Newmeyer, Frederick J. (2003). Grammar is grammar and usage is usage. *Language* 79, 682-707.
- Nooteboom, Sieb G. (1981). Speech rate and segmental perception, or the role of words in phoneme identification. In Myers, Terry, John Laver & John Anderson (eds.) *The Cognitive Representation of Speech*. Amsterdam: North-Holland Publishing Company, 143-150.
- Ohala, John J. (1990). The phonetics and phonology of aspects of assimilation. In Kingston, John & Mary E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press, 258-275.
- Ohala, John J. (1993). The phonetics of sound change. In Jones, Charles (ed.) *Historical Linguistics: Problems and Perspectives*. London: Longman, 235-278.
- Paolillo, John C. (2002). *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford: CSLI.
- Pater, Joe. (2009). Weighted constraints in generative linguistics. *Cognitive Science* 33, 999-1035.
- Pierrehumbert, Janet. (2001). Exemplar dynamics: word frequency, lenition and contrast. In Bybee, Joan L. & Paul Hopper (eds.) *Frequency Effects and the Emergence of Lexical Structure*. Amsterdam: Benjamins, 137-157.
- Pitt, Mark A., Laura C. Dilley, Keith Johnson, S. Kieling, William D. Raymond, Elizabeth Hume & E. Fosler-Lussier. (2007). *Buckeye Corpus of Conversational Speech*. (2nd Release). [<http://www.buckeyecorpus.osu.edu>]. Columbus: Department of Psychology, Ohio State University.
- Pitt, Mark A. & Arthur G. Samuel. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance* 19, 699-725.
- Port, Robert F. (1977). The influence of tempo on stop closure duration as a cue for voicing and place. *Haskins Laboratories: Status report on Speech Research* 51/52, 59-73.
- Port, Robert F. & Johnathan M. Dalby. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32, 141-152.
- Postal, Paul. (1968). *Aspects of Phonological Theory*. New York: Harper & Row.
- Prince, Alan & Paul Smolensky. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms. New Brunswick, Rutgers University. [Revised version appeared as Prince & Smolensky 2004.].
- Prince, Alan & Paul Smolensky. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [Revised version of Prince & Smolensky 1993.].

- Raymond, William D., Robin Dautricourt & Elizabeth Hume. (2006). Word-medial /t,d/ deletion in spontaneous speech: modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18, 55-97.
- Reynolds, Bill (1994). *Variation and Phonological Theory*. Ph.D. Dissertation, University of Pennsylvania.
- Sankoff, David. (1987). Variable rules. In Ammon, Ulrich, Norbert Dittmar & Klaus J. Mattheier (eds.) *Sociolinguistics: An International Handbook of the Science of Language and Society. Volume 2*. Berlin: Walter de Gruyter, 984-997.
- Sankoff, David, Sali Tagliamonte & Eric Smith. (2005). *Goldvarb X: A Variable Rule Application for Macintosh and Windows*. Department of Linguistics, University of Toronto. [Software package available at [http://individual.utoronto.ca/tagliamonte/Goldvarb/GV\\_index.htm.%5D](http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.%5D)].
- Smolensky, Paul. (2006). Harmony in linguistic cognition. *Cognitive Science* 30, 779-801.
- Sonderegger, Morgan (2012). *Phonetic and Phonological Dynamics on Reality Television*. Ph.D., University of Chicago.
- Sonu, Mee, Takayuki Arai, Keiichi Tajima & Hiroaki Kato. (2013). Effect of speaking rate variation on the perception of singleton and geminate consonants in Japanese by native and Korean listeners. *Proceedings of Meetings on Acoustics, Acoustical Society of America* 19.
- Temple, Rosalind A.M. (2009). (t,d): the variable status of a variable rule. In Parker Jones, 'Ōiwi & Elinor Payne (eds.) *Oxford University Working Papers in Linguistics, Philology & Phonetics: Papers in Phonetics and Computational Linguistics . Volume 12*. Oxford: University of Oxford.
- van Oostendorp, Marc. (1997). Style levels in conflict resolution. In Hinskens, Frans, Roeland van Hout & Leo Wetzels (eds.) *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 207-229.
- Walker, James A. (2012). Form, function, and frequency in phonological variation. *Language Variation and Change* 24, 397-415.
- Wang, Marilyn D. & Robert C. Bilger. (1973). Consonant confusion in noise: a study of perceptual features. *Journal of the Acoustical Society of America* 54, 1248-1266.
- Zimmerer, Frank, Henning Reetz & Aditi Lahiri. (2009). Place assimilation across words in running speech: Corpus analysis and perception. *Journal of the Acoustical Society of America* 125, 2307-2322.
- Zoll, Cheryl (1996). *Parsing Below the Segment in a Constraint Based Framework*. Ph.D., University of California, Berkeley.