Phonology **33** (2016) 425–458. © Cambridge University Press 2017
doi:10.1017/S0952675716000208

Probabilistic underspecification in nasal place assimilation[*]

John Coleman
University of Oxford

Margaret E. L. Renwick
University of Georgia

Rosalind A. M. Temple
University of Oxford

According to many works on English phonology, word-final alveolar consonants – and *only* alveolar consonants – assimilate to following word-initial consonants, e.g. *ran quickly* → *ra*[ŋ] *quickly*. Some phonologists explain the readiness of alveolar consonants to assimilate (*vs*. the resistance of velar and labial articulations) by proposing that they have underspecified place of articulation (e.g. Avery & Rice 1989). Labial or dorsal nasals do not undergo assimilation because their PLACE nodes are specified. There *are* reports that velar and labial consonants sometimes assimilate in English, but these are anecdotal observations, with no available audio and no statistics on their occurrence. We find evidence of assimilation of labial and velar nasals in the Audio British National Corpus, motivating a new, quantitative phonological framework: a statistical model of underspecification and variation which captures typical as well as less common but systematic patterns seen in non-coronal assimilation.

1 Introduction

According to many handbooks and textbooks on English phonology (e.g. Kreidler 1989: 237, Harris 1994: 72, Roca & Johnson 1999: 94, McMahon 2002: 45, Shockey 2003: 18–19), word-final alveolar consonants

[*] E-mail: JOHN.COLEMAN@PHON.OX.AC.UK, MRENWICK@UGA.EDU, ROSALIND.TEMPLE@MOD-LANGS.OX.AC.UK.

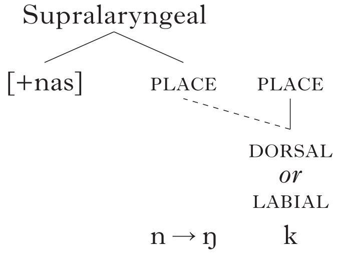
We thank our co-workers, Ladan Baghai-Ravary, Greg Kochanski and John Pybus, for their invaluable contributions to the creation, forced alignment and curation of the Audio BNC, without which the analyses presented here would not have been possible. We are grateful also to the referees, editors, colleagues, students and conference attendees who have read or heard earlier presentations of this work and have tested our claims and reasoning at each turn, forcing us to strive for maximum clarity. We claim all faults. Financial support for the research presented here was provided by the UK Economic and Social Science Research Council under award number RES-062-23-2566.

(i.e. /t d n s z/) – and notably *only* alveolar consonants – vary their place of articulation to match the consonant with which the next word begins, as in (1).

(1) that case	→ tha[k] case	cf.	*black case*
bad case	→ ba[g] case		*bag case*
ran quickly	→ ra[ŋ] quickly		*rang quickly*
this shop	→ thi[ʃ] shop		*fish shop*
his shop	→ hi[ʒ] shop		

Some phonologists have tried to explain the readiness of alveolar consonants to assimilate (*vs.* the resistance of velar and labial articulations to assimilation) by proposing that alveolar consonants are UNDERSPECIFIED for (i.e. they lack) place of articulation features (e.g. Avery & Rice 1989). Labial and dorsal places of articulation, in contrast, are specified by overt features, which, following Kiparsky (1985: 99), may spread backwards into a preceding empty (i.e. alveolar) PLACE node, according to a rule or constraint[1] like (2).

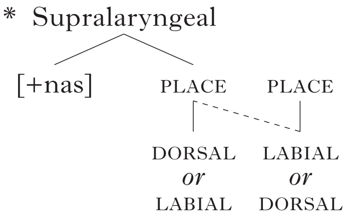
(2)



e.g. ra[ŋ] quickly

On this proposal, LABIAL (e.g. bilabial) or DORSAL (e.g. velar) nasals do not undergo such assimilation, because their PLACE nodes already contain features. This can be formalised with a constraint such as (3), prohibiting a LABIAL or DORSAL place feature from spreading backwards into a PLACE node that already bears a place feature.

(3)



[1] For this paper, it does not matter which: (2) can be interpreted as an (optional) rule stating that the specified change is permitted, or equally as a well-formedness constraint positively licensing the sharing of LABIAL or DORSAL place across the two PLACE nodes. Bird (1992) formalises the well-formedness constraint as a prohibition against a sequence of non-identical PLACE specifications.

According to (3), labial or velar consonants should not undergo assimilation; pronunciations such as *ki*[m]*pin* for *kingpin* or *alar*[ŋ] *clock* for *alarm clock* should not occur. Nevertheless there *are* in fact reports that velar and labial final consonants sometimes assimilate in English, as in (4).

(4)	like that	→	li[t]e that	(Barry 1985)
	from Kingston	→	fro[ŋ] Kingston	(Avery & Rice 1989)
	I'm going	→	I'[ŋ] going	(Ogden 1999)
	some girls	→	so[ŋ]e girls	(Lodge 2009)
	same night	→	sa[n]e night	(Cruttenden 2014)
	same kind	→	sa[ŋ]e kind	(Cruttenden 2014)
	King Charles	→	Ki[n] Charles	(Cruttenden 2014)

However, these are anecdotal observations, with no context, no audio available for detailed study and no statistics on their frequency of occurrence. It is conceivable that such sporadic counterexamples could be speech errors, dysfluencies or other kinds of pathological forms that do not reflect normal phonological competence and do not warrant radical revision to phonological theory. Alternatively, such examples could be well-formed in some varieties of English but not others, or they could be optional but relatively uncommon. It is crucial to study the systematicity of potential counterexamples in order to understand the nature of phonological specification.

The increasing number and size of speech corpora and advances in speech technology now provide unprecedented opportunities to study large quantities of real-life speech, in order to answer linguistic questions which it has not previously been possible to address in smaller-scale studies. Large corpora allow the investigation of phenomena which are systematic, and may therefore be relevant for modelling phonological processes, but which are also rare, and thus have previously lacked adequate empirical investigation. Assimilation of word-final alveolars to following consonants has been studied in an American English corpus by Dilley & Pitt (2007), but that paper did not look for instances of bilabial or velar assimilation. We suspect that assimilation in non-alveolar nasals has been previously overlooked in the literature precisely because it is relatively rare. Therefore, a large-scale analysis is called for, based on thousands of tokens from spontaneous speech, to demonstrate that assimilation of word-final labial and velar consonants does indeed occur.

In this paper, we take advantage of the spontaneity and size of the Audio BNC (British National Corpus), one of the largest archives of fully transcribed 'language in the wild' ever collected, containing almost 1500 hours of recorded speech, to assess the occurrence of labial and velar nasal assimilations, focusing on pairs of words where the first ends in a nasal consonant and the second begins with a non-nasal consonant. Spontaneity is important because velar or labial assimilations may be far less likely to occur in careful laboratory speech. Large size is needed because of the rarity of non-canonical assimilation, and because of the extremely unbalanced distributions of linguistic units (phonemes, syntactic constructions, words) in natural language – by Zipf's Law (e.g. Zipf 1935, Mandelbrot 1961, Miller & Chomsky 1963), some sounds, words, pairs of words, etc. are vastly more frequent than others.

In addition to their unbalanced nature, recordings of spontaneous speech are often noisy. To show the presence of multiple phonetic realisations within a dataset, such as a mixture of assimilated and unassimilated nasals, subsets of the data must be demonstrably different from one another, with statistically significant differences along some phonetic parameter(s). The number of tokens needed to attain statistical significance is primarily a function of the amount of variation in the data.[2] For speech recorded from a single speaker in good conditions, the variance may be relatively low, and therefore a small number of tokens might be adequate; for speech recorded from many speakers and varieties, as in a naturalistic corpus, a much larger number of tokens may be needed in order to make statistically valid inferences. To illustrate this, consider the histograms of F2 frequency measurements in Fig. 1, which show relative proportions of subsets of our data falling into 100 Hz frequency bins. For the rather 'noisy' /ŋ/ data presented in (a) to be a statistically representative sample, with a confidence level of $p < 0.05$, a measurement error E of up to 100 Hz (the size of the histogram bins), and an empirically estimated standard deviation σ = 371 Hz, the number of tokens, N, needs to be at least 53. Even in such a large dataset as the Audio BNC, we find only 33 tokens of /ŋ/ in this context from female speakers, which is statistically insufficient. For the /m/ data in (b), to attain a higher confidence level of $p < 0.01$, with a standard deviation estimate of 250 Hz and a measurement error of only 50 Hz, N must be at least 167. The 736 tokens available are in this case sufficient, because the variance is lower. Figure 1 thus illustrates the need to have a sufficient number of tokens in order to obtain a statistically well-behaved distribution, i.e. one with a clearly defined central tendency. As a rule of thumb, therefore, we aim to find hundreds rather than tens of tokens of items of interest. The size of the Audio BNC is thus crucial in allowing us to overcome variation.

In the rest of this paper we lay out evidence for labial and velar nasal assimilations in British English, and present an alternative model of assimilation. We restrict the study to nasals, so that we can measure formant frequencies during their closure portion, which is not possible with oral stops. §2 describes how we prepared the corpus and harvested the relevant word-pairs before outlining our methods. In §3 we present both qualitative and quantitative data to show that non-alveolar nasal assimilation *does* in fact occur in English, albeit not as frequently as with coronals. Since this means that nasals at *all* places of articulation may assimilate, a new approach to the phonology of nasal place of articulation variation is required; we

[2] Specifically, $N = (z \times \sigma/E)^2$, where z is the level of confidence desired (z = 0.95 for $p < 0.05$, or z = 0.98 for $p < 0.01$), σ is the standard deviation in the population (of which the dataset is taken to be a representative sample) and E is the maximum allowable error.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

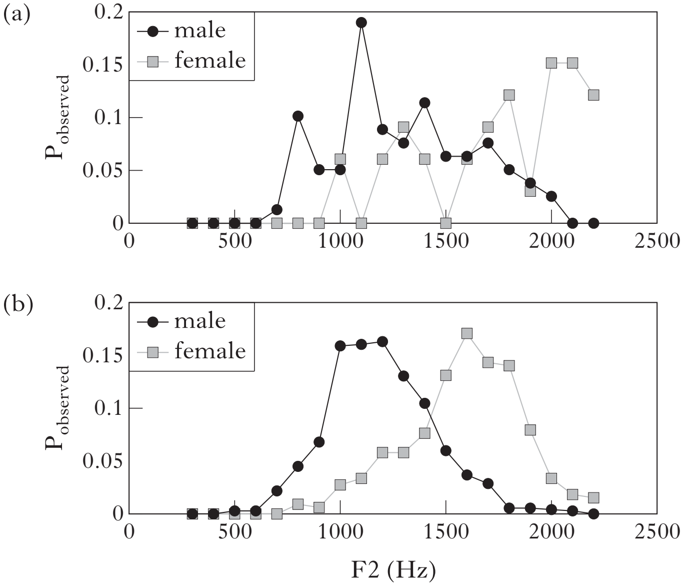


Figure 1

Illustration of how larger sample sizes can yield smoother, single-peaked, distributions. (a) Histograms of F2 of /ŋ/ before /k/ or /g/ (male speakers: $N = 79$, mean F2 (μ) = 1299 Hz, standard deviation (σ) = 336 Hz; female speakers: $N = 33$, $\mu = 1752$ Hz, $\sigma = 371$ Hz); (b) histograms of F2 of /m/ in *from the* (male speakers: $N = 736$, $\mu = 1214$ Hz, $\sigma = 218$ Hz; female speakers: $N = 328$, $\mu = 1554$ Hz, $\sigma = 302$ Hz). The vertical axis is relative incidence in the corpus sample (count/N).

present in §4 a new, probabilistic framework for modelling phonological variation – a quantitative underspecification theory – and show how data on assimilated and unassimilated forms can be modelled in that framework. This new approach demonstrates how probabilistic gradience, categories and underspecification may be reconciled.

2 Methods and procedures

2.1 Preparing the Audio British National Corpus

Collected in 1991–92, the 10-million-word Audio British National Corpus was designed to include speech from across the United Kingdom (Crowdy 1993, Coleman *et al.* 2012). Roughly half the corpus consists of unstructured, informal speech collected by volunteers, while the other half is largely unscripted speech collected in more formal settings, such as interviews and religious services. This spoken material was originally recorded on 1213 cassette tapes, which were transcribed orthographically by professional audio typists. The corpus was originally published only as

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

linguistically annotated orthographic transcriptions, together with speaker-specific metadata about age, sex, occupation and location, and other details of sociolinguistic relevance such as the relationship of the speaker to the volunteer who made the recordings (Crowdy 1995), as part of the British National Corpus (BNC Consortium 2007). In 2009–10, the British Library Sound Archive digitised most of the original audio recordings (6.9 million words of spoken audio) to stereo PCM audio (.wav files) at 96 kHz with 24-bit resolution.

The first major challenge in mining an audio corpus for a sufficient number of examples of the phenomenon being studied is the task of simply locating the relevant tokens. The almost 1500 hours (or two months) of continuous audio contains tens of thousands of instances of word-final nasals before word-initial consonants. It is hardly feasible to locate examples just by listening to the recordings and manually marking or editing the relevant instances; an automatic method of aligning a transcription with the audio is necessary.

We downsampled the high-resolution recordings to 16-bit, 16 kHz monophonic audio files. Alignment of the transcriptions to the audio files was performed automatically, using the HTK speech-recognition toolkit (Young *et al*. 2009), with an HMM topology to match the Penn Phonetics Laboratory Forced Aligner (P2FA; Yuan & Liberman 2008). Our alignment system used acoustic models that combined the P2FA American English models with our own British English models, to provide acoustic matching with a wide range of possible pronunciations (Baghai-Ravary *et al*. 2011). In accordance with the recording agreements and publication principles of the BNC transcriptions (Crowdy 1994), personal names and some other speaker-specific information in the recordings were silenced to respect speaker anonymity. The alignment output includes Praat TextGrids (Boersma & Weenink 2012), whose tiers contain segment- and word-level transcriptions, as in Fig. 2 below.

As this study examines word-final nasals in various following contexts, we first surveyed all pairwise combinations of words occurring in the corpus where the first word ends with a nasal, for example *some cream*. From the Praat TextGrids generated by forced alignment, we compiled an index of word-pair locations (filename, word-pair start and end times) for the entire corpus. Speaker metadata are available with varying level of detail for 64.8% of the 6.9 million word-pairs in the corpus. These were merged with acoustic alignment information to create a single index of lexical, segmental, timing and socio-indexical data. Although the automatic alignment system was highly trained, it was not error-free. Thus, once word-pairs of interest (see §2.2) had been found in the index, every token was listened to in order to exclude from further analysis all tokens that had been grossly misaligned. Approximately 37,000 tokens were checked in this way, with transcription and audio audibly aligned in 67% of cases; in one-third of entries identified by the index, the complete word-pair was not audible in the corresponding audio clip. Misaligned audio clips were removed from the dataset. From the verified word-pairs, the analysis was

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

further restricted to tokens for which the speaker's sex was noted in the corpus metadata, so that we could employ different settings for female and male voices when measuring formant frequencies (see §2.3 below).

Word-pairs selected for analysis were then extracted from the original audio files and realigned automatically, using a modified dictionary that listed all potential assimilated pronunciations. The purpose of this realignment was to improve segment-boundary locations by allowing for shorter segment durations and the possibility of alternative segment labels: an automatic aligner performs best on small portions of speech for which the orthographic transcription is known precisely, while longer stretches of audio (such as the original tape-recordings) require more processing time and are prone to greater error.

The recordings in this corpus are quite challenging for forced alignment and formant-frequency tracking. Due to the informal recording methods (Sony Walkman cassette recorders with built-in condenser microphones, used by volunteer members of the public in a wide variety of recording environments), the signal-to-noise ratio in many of the recordings is so poor that it can be extremely difficult even for an expert to discern formants or cues to segment boundaries using visual examination of their spectrograms. Therefore, we evaluated the accuracy of word and segment boundaries assigned by the forced aligner against two reference sets of hand-corrected boundaries. For the word-boundary evaluation, the absolute differences between automatic and manually corrected times at three data points (the start and end of word 1, and the end of word 2) were calculated for 549 tokens of the highest-frequency word-pairs. 60% of the automatically assigned boundaries fell within 50 ms of the corresponding manual boundaries and 80% within 100 ms; the root mean square (RMS) difference was 70 ms. For the segment-boundary evaluation, the start and end times of the word-final nasals in 374 tokens of *come back*, 126 tokens of *coming back* and 99 tokens of *coming down* were examined. 50% of the automatically assigned boundaries were within 50 ms of the corresponding manual boundaries, 65% within 70 ms and 80% within 100 ms; the RMS difference was 80 ms. Crucially, these differences had no material effect on statistical analyses presented in §3 below, giving us confidence in the validity of using automatically aligned data. Consequently, the measurements and statistics reported below are based on tokens that were located using the automatic alignments.

2.2 Materials: word-pair selection

To identify environments for potential non-canonical assimilation, we searched the word-pair index for word-pairs in which the first word ends in a nasal consonant, and the second begins in an oral consonant. We searched for bilabial nasal /m/ before a velar stop (e.g. *I'm getting*), and before an alveolar (e.g. *I'm trying*), as well as in non-assimilation control contexts (before another bilabial, e.g. *I'm putting*). Likewise, we searched for velar nasal /ŋ/ before an alveolar stop (e.g. *long time, trying to*) or a bilabial (e.g. *young people, coming back*), and in a velar–velar

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

control context (e.g. *something called, dressing gown*). Gerunds such as *coming* may also have regular [n] variants (i.e. *comin'*), as reported in the sociolinguistics literature (e.g. Trudgill 1974, Labov 1989). Because bilabial tokens might thus be regarded as assimilated forms of these alveolar variants (Yuan & Liberman 2011), the interpretation of evidence of assimilation in these verb forms is somewhat problematic. That is, any assimilation of *-ing* that we may observe could be assimilation of the coronal nasal allomorph to the following consonant, rather than a counterexample to coronal underspecification. Therefore, we were careful to select a number of non-gerunds with final /ŋ/, in which such '*g*-dropping' is not expected. We also located pairs in which the first word ends in an alveolar /n/, before a labial or velar stop – canonical assimilation contexts – as controls. Examples with at least ten instances are listed in Table I, together

	initial /p, b/	initial /ð/	initial /t, d/	initial /k, g/

final /m/	some people 166
	I'm putting 10
	come back 370
	them back 88
	them but 45
	him but 30
	him back 28
		from the 1064
		from there 127
		them that 79
			seem to 242
			I'm trying 72
			I'm talking 45
			come down 321
			I'm doing 106
			them down 86
			them do 26
				I'm getting 58
				I'm glad 50
				I'm going 414
				I'm gonna 318
				some cream 8

final /n/	in particular 32 in bed 97 in between 74 in Britain 33	on the 4606 on there 239 than that 123	in to 136 in time 25 nine ten 15 *seen to* 12 *been doing* 104 then do 36	in case 224 can come 121 on come 120 European Community 33 then go 54 been going 53
final /ŋ/	young people 52 swimming pool 22 *coming back* 125 going back 110 *nothing but* 57 *something but* 28	doing that 163 saying that 198 thing that 163	trying to 668 talking to 206 long time 174 going down 127 *coming down* 98 *something different* 30	Hong Kong 21 something called 17 training course 11 dressing gown 13 young girl 12

Table I

Selection of the word-pairs analysed in this study, with the number of tokens that are well aligned and for which the speaker sex is recorded. Shaded cells contain non-assimilation control conditions, and cells with heavy borders are cases in which assimilation of /m/ or /ŋ/ to a following consonant might occur. Items discussed in detail below are italicised.

with some relevant word-pairs which occur less frequently, but which we shall discuss further below, e.g. *some cream*.

2.3 Formant-frequency analysis

We analysed formant frequencies in the word-final nasals and (to begin with) the vowels immediately preceding them in order to see whether there were any acoustic differences of the kind that are characteristic of place-of-articulation differences.[3] It is far from easy to analyse place of articulation of nasals from acoustic spectral data alone. Prior work (Fujimura 1962, Kurowski & Blumstein 1987, Stevens 1998) has shown that in nasal stops, the first formant of flanking vowels divides into two nasal resonances plus an antiresonance, a reduction in spectral energy corresponding to the resonance of the closed mouth cavity. The frequency of this antiresonance correlates well with place of articulation, being low for [m] (ca. 750–1250 Hz), higher for [n] (ca. 1450–2200 Hz) and highest, but more variable, for [ŋ]. However, automatic analysis tools typically only provide measurements of resonances (formants), not antiresonances, so we focused on measurements of formant frequencies during the nasal murmur.

Dilley & Pitt (2007) measured formant frequencies during the transition from the vowel into following consonants, in order to assess place of articulation. Since they were examining /t/ and /d/ as well as /n/, this method is appropriate for their study; however, measuring such vowel–nasal transitions is neither necessary nor very suitable for our study. Dilley & Pitt used the smaller Buckeye Corpus, which contains high-quality recordings of unscripted interviews that were manually labelled. The poorer audio quality of recordings in the BNC makes it more difficult to be confident about measurements of formant frequencies in a vowel-to-consonant transition than in the nasal consonant: identifying the correct time interval for measuring transitions requires accurate timing information about the segment boundaries; since the speech rate is very variable across the corpus, the duration of transition portions is very variable from one token to another. However, it is not necessary to look at vowel transitions in order to examine place differences, because it is possible and sufficient to examine the formant frequencies of the nasals themselves; we can examine the segment of interest directly. This is consistent with how listeners perceive place of articulation of postvocalic nasals: Repp & Svastikula (1988) found that listeners made use of the spectrum of the nasal murmur itself (in addition to the quality of the preceding vowels), whereas formant movements in the vowel–nasal transition were *not* perceptually salient, unlike the case with nasal–vowel transitions. From our previous experience of measuring acoustic cues to place of articulation (e.g. Olive *et al*. 1993), we expected that F1 would not be very different for [m], [n] and [ŋ]. We

[3] Besides total assimilation, other possible sources for apparent assimilation include coarticulation or gestural overlap (double articulation) between a word-final nasal and a following word-initial consonant. We discuss these possibilities below.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

expected that [m] would have the lowest F2 and F3, and that [ŋ] would have a higher F2, possibly rising in the direction of a falling F3, especially after front vowels (the 'velar/palatal pinch'). In front vowel contexts especially, F2 of [ŋ] could be very high, and possibly even higher than F2 of [n]. In general, however, we expected that F2 and F3 for [n] would be highest of all, modulo the variation due to assimilation that has been well documented for [n]. These patterns are also evident in the qualitative descriptions and figures in Potter *et al.* (1966: 189).

Using Praat, we automatically measured formant frequencies in the aligned word-pairs. Each token was downsampled to 8 kHz, and formant frequencies were measured via a short-term spectral analysis with a window length of 25 ms and 50 Hz pre-emphasis. Data from male and female speakers were measured and analysed separately. For male speakers, five formants were measured, with a maximum range of 4500 Hz, while for female speakers, four formants were measured, with a maximum range of 5500 Hz. In order to normalise over segments of different durations, the formant frequencies were measured at 10% fractions of each segment in a word-pair (from 0% to 100% of the segment's duration). As the formant frequencies are quite stable during the nasals examined in this paper, we averaged across all deciles to obtain a single mean frequency for each formant, for each nasal token.

3 Results

We looked for evidence of assimilation of word-final /m/ or /ŋ/ to following consonants in four complementary ways. First, we examined relevant audio portions impressionistically, i.e. by listening to many such word-pairs and inspecting their spectrograms (§3.1). This part of the study includes some word-pairs for which only a few tokens are available, and which are therefore not amenable to statistical analysis. However, they may be considered as 'existence proofs' of non-coronal assimilation. Then, using linear mixed-effects models (Baayen *et al.* 2008), we tested whether the place of articulation of following consonants has a statistically significant effect on the F2 frequency of final nasals (§3.2). Finally, we examined the statistical distributions of F2 frequency variation using histograms (§3.3), and making planned comparisons to see how word-final nasals in certain, similar contexts were or were not differentiated (§3.4). As few significant differences in formant frequencies were found for F1 or F3, the analyses and results given here focus on F2.[4]

3.1 Impressionistic evidence of non-coronal assimilation

In the impressionistic part of this study, one of the authors, an experienced ear-trained phonetician, listened to 668 tokens that were candidate cases of

[4] We report the F1 and F3 measurements in an earlier conference paper (Renwick *et al.* 2013). In that paper we also present data on a further analysis, in which we allowed the forced alignment programme to decide whether to label word-final nasals as M (i.e. [m]), N (i.e. [n]) or NG (i.e. [ŋ]), according to the signal acoustics.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

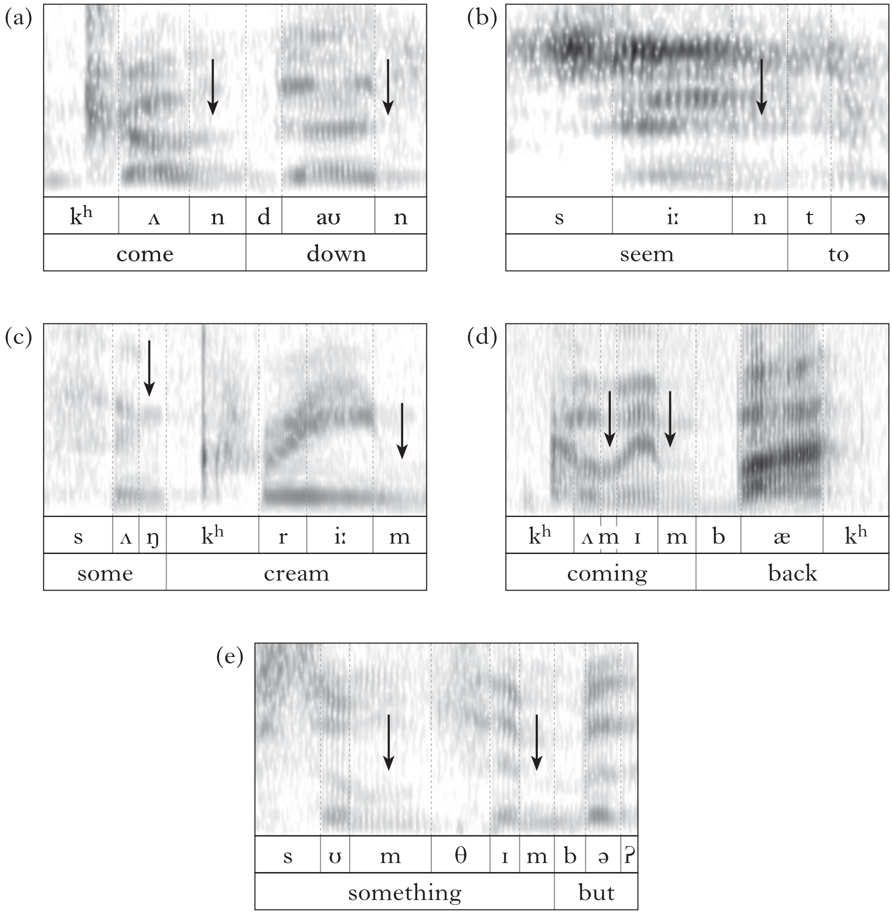


Figure 2

Wide-band spectrograms of (a) *come down*, with final nasal in *come* pronounced [n] (cf. [n] in *down*); (b) *seem to*, with final nasal in *seem* pronounced [n] (cf. [n] in *down*); (c) *some cream*, with final nasal in *some* pronounced [ŋ] (cf. [m] in *cream*); (d) *coming back*, with final nasal in *coming* pronounced [m] (cf. [m] in *some*-); (e) *something but*, with final nasal in *something* pronounced [m] (cf. [m] in *some*-). The transcriptions were added manually, not automatically. Arrows indicate the F2 of nasals.

bilabial or dorsal assimilation. A number of clear cases of assimilated non-alveolar nasals were identified and then examined in more detail in spectrograms. Figure 2 presents spectrograms of five examples of these audibly assimilated non-alveolar nasals at word boundaries. In Fig. 2a, the F2 of /m/ in *come down* does not drop, as it does in the /m/ of *cream* (Fig. 2c) and *coming* (Fig. 2d), but remains high, similar to the /n/ in *down*. Likewise, in Fig. 2b, the word-final /m/ in *seem to* has a high F2 that remains high during the short [t] closure that follows. In Fig. 2c, the F2 of /m/ in *some* clearly rises towards F3, forming a 'velar pinch' before the /k/ of *cream*, an unmistakable characteristic of [ŋ] (Olive *et al*. 1993: 97). In Fig. 2d, the /ŋ/ in *coming back* has a low F2, similar to that of the medial /m/ of *coming* and the final /m/ of *cream* in Fig. 2c. Again, in Fig. 2e, the final /ŋ/ in *something but* has a low F2, similar to the [m] tokens in Fig. 2d, but quite unlike the 'velar pinch' evident in Fig. 2c.

As expected, such clearly audible assimilations are relatively uncommon: whereas 20% of alveolar nasals were judged to be assimilated in Dilley & Pitt (2007), in this data 8.3% (18/216) of lexically velar nasals and 4% (37/976) of bilabial nasals were judged to be assimilated.[5] However, the subjective listening test has limitations, such as listener biases and difficulties with unclear or ambiguous stimuli. Although phonetic listening and the inspection of spectrograms were useful for revealing the existence of a non-negligible number of examples of non-coronal assimilation, including word-pairs such as *alar*[ŋ] *clock*, where the number of tokens is small, statistical tests are required to assess the systematicity of non-coronal assimilation.

3.2 Linear mixed-effects models of F2 frequency in nasals

If nasals in word-final position assimilate systematically to following consonants, we should find acoustic variation in those nasals consistent with variation in place of articulation: nasals before labials should show a reduction in mean F2 frequency, nasals before alveolars should show relatively high F2 frequency and nasals before velars should have F2 frequencies intermediate between pre-labial and pre-alveolar nasals. To test this, we first computed the average F2 frequency of each word-final nasal in 14,402 tokens of 444 word-pairs in control and assimilation contexts. There are many sources of noise in the data, especially speaker-to-speaker variation (1227 speakers) and the quality of the pre-nasal vowel. We control for these confounds by fitting linear mixed-effects models to the data. This type of model allows us to test the fixed factors expected to have a constant effect on nasal formant frequencies, while simultaneously accounting for variation across speakers and word-pairs, and maintaining robustness in the presence of unequal numbers of observations. The analysis was run in R using the lme4 package (Bates & Maechler 2009). Linear mixed-effects models were fitted by hand, using model comparison, and *p*-values were obtained using the lmerTest package (Kuznetsova *et al*. 2013). A model was fitted for each lexical nasal place of articulation; the dependent variable was F2 frequency, with one

[5] 5/30 tokens of *something different* sounded alveolar, while 6/89 tokens of *something but*, 6/92 tokens of *nothing but* and 1/5 tokens of *wrong* before a /p/ or /b/ sounded labial. 14/490 tokens of *seem to*, 3/109 tokens of *come down* and 2/34 tokens of *some* followed by /t/ or /d/ were heard as alveolar; before a dental, 8/215 tokens of *from the* sounded unambiguously coronal, e.g. [fɹən̪n̪ə] (see Manuel 1995), while 2/53 tokens of *from there* were audibly assimilated. 2/19 tokens of *alarm clock* and 6/56 tokens of *some cream* sounded clearly velar.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

(average) measurement for each token. A random intercept was fitted for speaker. We included fixed effects for sex, following place of articulation and preceding vowel quality.

Summaries of the best-fitting models for /m, n, ŋ/ are shown in Tables II–IV, where the β-value estimates may be interpreted as the amount (in Hz) by which the average F2 frequency of the nasal is raised or lowered by a given factor level. In all three models, the fixed effects were significant predictors of nasal F2 frequency. Unsurprisingly, speaker sex and preceding vowel have an effect on F2: F2 frequencies are lower for male than for female speakers, and front vowels induce higher nasal F2 frequencies than non-front vowels. In each of the three models we also find significant effects of following consonant place of articulation. In each model the F2 frequency is significantly higher before alveolars or velars than before labials. Although there is little difference between the effect of following alveolars and velars, the fact that all three nasals vary significantly according to the following consonant place confirms that assimilation is not limited to coronals.

To further explore the implications of these findings, we now examine the distributional patterns of the F2 frequency variation more closely, and carry out planned comparisons between selected word-pairs.

	β estimate	standard error	t	p
Intercept	1625.012	15.309	106.145	~0
Sex = male	−344.590	12.870	−26.774	~0
Following Place = alveolar	30.243	10.605	2.852	<0.01
Following Place = velar	34.749	13.511	2.572	<0.05
Preceding V = [ˈʌ]	−36.954	11.544	−3.201	<0.005
Preceding V = [ˈɔ]	−158.166	24.653	−6.416	~0
Preceding V = [ˈɑɪ]	41.596	20.170	2.062	<0.05
Preceding V = [ˈɛ]	121.555	26.085	4.660	<0.0001
Preceding V = [ɚ]	6.068	13.758	0.441	0.659
Preceding V = [ˈɪ]	212.001	42.919	4.940	~0
Preceding V = [ˌɪ]	158.672	48.372	3.280	<0.005
Preceding V = [ˈi]	223.865	18.624	12.020	~0
Preceding V = [ˈeɪ]	193.728	17.199	11.264	~0

Table II

Summary of best mixed-effects model for /m/ (N = 4441, 870 speakers). The reference level for this model is the control condition, i.e. /m/ followed by another labial consonant, with Sex = female and Preceding V = [ə]. Random factor = (1 | Speaker).

	β estimate	standard error	*t*	*p*
Intercept	1559.735	9.920	157.226	~0
Sex = male	−326.774	12.256	−26.663	~0
Following Place = labial	−24.585	10.451	−2.352	<0.05
Following Place = velar	16.374	14.042	1.166	0.244
Preceding V = [ˈa]	94.295	49.745	1.896	0.058
Preceding V = [ə]	224.523	11.372	19.743	~0
Preceding V = [ˈɔ]	−21.849	8.218	−2.659	<0.01
Preceding V = [ˈɑɪ]	208.171	60.206	3.458	<0.001
Preceding V = [ˈɛ]	270.546	16.379	16.517	~0
Preceding V = [ˈɪ]	303.978	12.955	23.465	~0
Preceding V = [ˈi]	268.045	24.294	11.033	~0

Table III

Summary of best mixed-effects model for /n/ (*N* = 6991, 987 speakers). The reference level for this model is the control condition, i.e. /n/ followed by another alveolar consonant, with Sex = female and Preceding V = [ˈɑ]. Random factor = (1 | Speaker).

	β estimate	standard error	*t*	*p*
Intercept	1722.494	32.121	53.626	~0
Sex = male	−347.663	14.909	−23.319	~0
Following Place = labial	−54.880	24.417	−2.248	<0.05
Following Place = alveolar	3.443	22.282	0.155	0.877
Preceding V = [ˈʌ]	−119.389	33.198	−3.596	<0.001
Preceding V = [ˈɔ]	−215.420	24.589	−8.761	~0
Preceding V = [ˈɑɪ]	41.304	24.245	1.704	0.089
Preceding V = [ɪ]	99.175	22.939	4.323	<0.0001
Preceding V = [ˈɪ]	165.337	23.881	6.923	~0

Table IV

Summary of best mixed-effects model for /ŋ/ (*N* = 2970, 768 speakers). The reference level for this model is the control condition, i.e. /ŋ/ followed by another velar consonant, with Sex = female and Preceding V = [ə]. Random factor = (1 | Speaker).

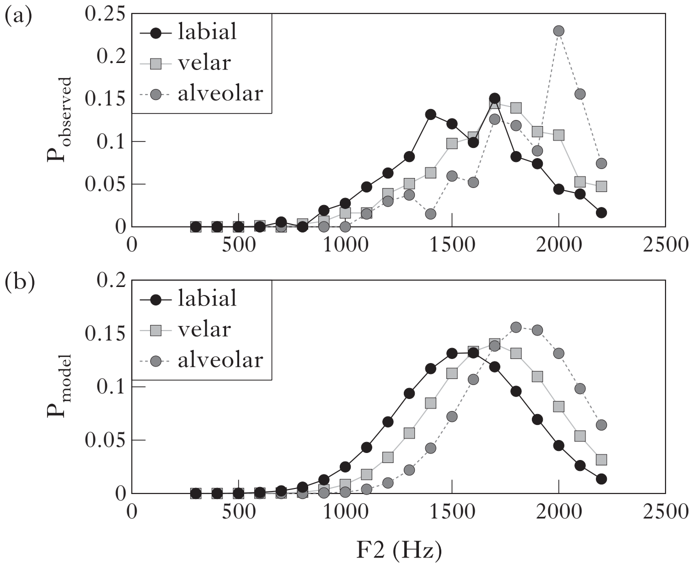


Figure 3

(a) Histograms of F2 of nasals (female speakers). Labial: /m/ before /p, b/ (N = 365); velar: /ŋ/ in all contexts (N = 933); alveolar: /n/ before /t, d/ (N = 135). (b) Gaussian standard normal distributions (probability density functions) fitted to those histograms. Labial: model of /m/ (a = 36,088, μ = 1554 Hz, σ = 302 Hz); velar: model of /ŋ/ (a = 86,742, μ = 1694 Hz, σ = 293 Hz); alveolar: model of /n/ (a = 10,067, μ = 1838 Hz, σ = 271 Hz).

3.3 Controls: distributions of F2 variation in unassimilated nasals

There is surprisingly little published data on the acoustics of nasal consonants: earlier studies (e.g. Fujimura 1962, Kurowski & Blumstein 1987) typically report measurements of one or a few speakers. We therefore begin by examining the F2 frequency differences between /m, n, ŋ/ in non-assimilating environments (the shaded cells in Table I) in 11,669 tokens of 74 word-pairs spoken by 1181 speakers drawn from across the corpus. Figures 3a and 4a plot histograms of F2 variation in /m, n, ŋ/ produced by female and male speakers. While the data for /m/ and /n/ are drawn from pre-labial and pre-alveolar contexts respectively, there are too few data for /ŋ/ before /k/ or /g/ to generate smooth histograms, as demonstrated in Fig. 1 above. The F2 frequencies for [ŋ] are thus pooled across all following consonantal contexts, including a small proportion of potential assimilation contexts. As the actual incidence of assimilated tokens is very small (see below), this does not affect the overall F2 distribution a great deal.

In Table V, the means and standard deviations of F2 frequency are given for each category. These values were used to fit standard normal distributions (Gaussian probability density functions) to the histograms using the

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

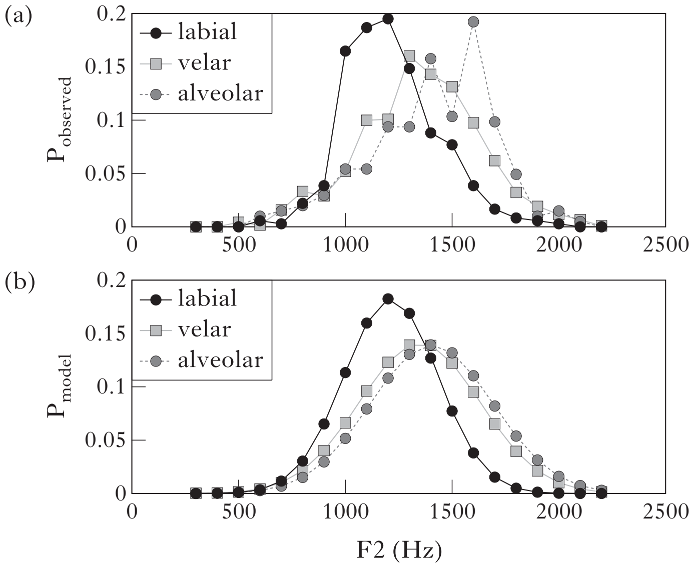


Figure 4

(a) Histograms of F2 of nasals (male speakers). Labial: /m/ before /p, b/ ($N = 364$); velar: /ŋ/ in all contexts ($N = 1211$); alveolar: /n/ before /t, d/ ($N = 203$). (b) Gaussian standard normal distributions (probability density functions) fitted to those histograms. Labial: model of /m/ ($a = 39{,}967$, $\mu = 1214$ Hz, $\sigma = 218$ Hz); velar: model of /ŋ/ ($a = 12{,}231$, $\mu = 1349$ Hz, $\sigma = 283$ Hz); alveolar: model of /n/, $a = 20{,}367$, $\mu = 1404$ Hz, $\sigma = 287$ Hz).

Matlab normpdf function,[6] with observed mean F2 frequency μ, the standard deviation from the mean σ and a scale constant a derived as a linguistically irrelevant by-product of the data-fitting procedure. These normal distributions are plotted in the lower panel of Figs 3b and 4b.

	female		male	
	mean F2	SD	mean F2	SD
/m/	1554	302	1214	218
/n/	1694	293	1349	283
/ŋ/	1838	271	1404	287

Table V

Means and standard deviations (in Hz) of F2 frequencies of canonical (unassimilated) /m, n, ŋ/ in non-assimilating environments across a wide range of speakers.

[6] http://uk.mathworks.com/help/stats/normpdf.html.

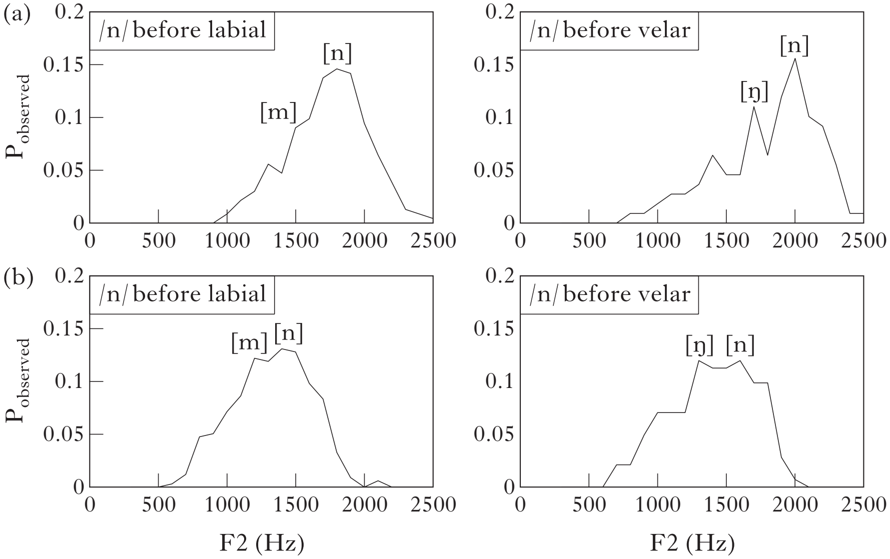


Figure 5

Histograms of F2 of word-final /n/ followed by bilabials and velars. (a) Female speakers (N= 233 and 109); (b) male speakers (N= 336 and 142).

In the observed distributions of F2 frequency seen in Figs 3a and 4a, a wide range of variation is evident within each category. This is because of the very wide range of speakers in the corpus. Even so, it can be seen for both males and females that the mode of F2 for /m/ is generally lowest, that of /n/ is highest and /ŋ/ in between.

Having established the population means and statistical distributions of F2 frequencies in unassimilated controls, we now examine whether nasals vary systematically according to the place of articulation of the following consonant.

3.4 Distributions of F2 variation in /n/-assimilation

Since /n/ is known to assimilate systematically to the place of articulation of following consonants, we here present histograms of F2 frequency in tokens of word-final /n/ in our corpus which are followed by bilabial and velar consonants. As can be seen in Fig. 5, assimilation gives rise to statistical distributions that are multimodal. Sometimes the second mode is clearly visible as a second peak, and sometimes it is not so clearly differentiated from the primary peak, appearing more as an elbow on the tail of the primary peak than as a separate peak. In all the histograms, the peak representing unassimilated [n] tokens falls at roughly the same frequency as in Figs 3 and 4 above (around 1800 to 2000 Hz for females and around 1400 to 1600 Hz for males). The secondary peaks representing tokens assimilated to following bilabials (labelled [m]) fall close to the F2 frequency

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

peaks for word-final /m/ in non-assimilation contexts in Figs 3 and 4 (around 1450 Hz and 1200 Hz). In the same way, the peaks representing tokens assimilated to velars (labelled /ŋ/) fall at F2 frequencies in between, at around 1700 Hz and 1300 Hz for female and male speakers respectively. Thus in canonical assimilation contexts we find a hierarchy of distributional peaks at F2 frequencies parallel to the /n/ > /ŋ/ > /m/ hierarchy found in our unassimilated control tokens. Therefore, if word-final bilabials and velars also assimilate systematically to following consonants, we expect to find similar patterns, albeit with lower-probability secondary peaks, given the lower rates of assimilation found in our auditory study (§3.1 above). To further explore the implications of these findings, we now analyse the distributional patterns of F2 frequencies in assimilation contexts, and carry out planned comparisons to examine the magnitude and significance of F2 differences between selected word-pairs.

3.5 Assimilation in /m/ and /ŋ/: planned comparisons

The distributions of word-pairs in the corpus are extremely lopsided, as illustrated in Table I. Moreover, the most frequent word-pairs (e.g. *on the*) are also the most likely to be spoken more quickly and reduced in casual speech, and are hardest for the aligner to delimit accurately. On the other hand, the formant frequencies of less frequent word-pairs (e.g. *young girl*) are highly variable across tokens, making statistical comparisons difficult. The clearest data come from word-pairs that are frequent enough for statistically useful results, but not so frequent as to be prone to extreme phonetic shortening. We therefore focus on a selection of such pairs in our planned comparisons.

3.5.1 *Variation in bilabial nasals.* Table VI displays the means and standard deviations of nasal formant frequencies in *seem to* and *been doing*.[7]

	sex	tokens	F2	SD
been doing /n/	female male	57 50	1917 1474	257 274
seem to /m/ ~ [n]	female male	105 147	1850 1526	349 258

Table VI

Means and standard deviations (in Hz) of F2 frequencies of nasals. ‘A ~ B’ indicates ‘A tending towards B’.

[7] We also examined the minimal pair *seem to vs. seen to*, but unfortunately there are not enough tokens of *seen to* in the corpus to support good measurements of this contrast.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

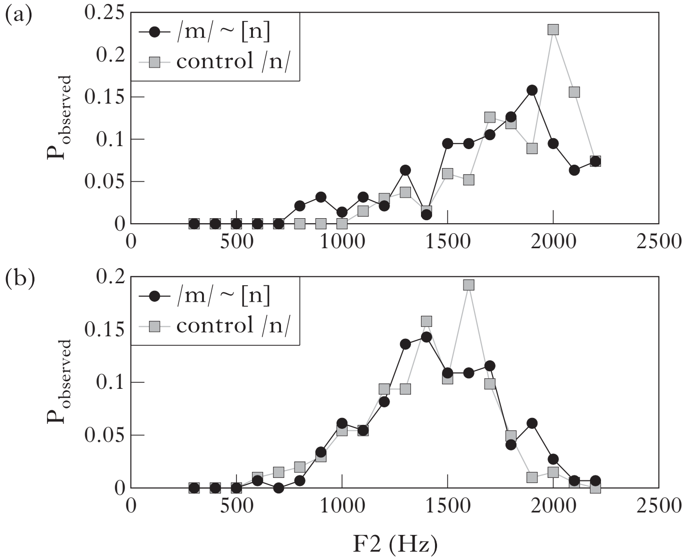


Figure 6

F2 variation in /m/ of *seem to vs.* control /n/. (a) Female speakers; (b) male speakers.

If the /m/ in *seem to* sometimes assimilates to the following /t/, it will have higher mean F2 frequency than before a labial consonant. Unfortunately, the corpus has too few controls of the form /iːm # {p, b}/, so we are unable to assess this comparison directly. If, however, the /m/ in *seem to* retains its canonical pronunciation, and never assimilates to the following /t/, we expect it to have a lower mean F2 than the /n/ in *been doing*, consistent with their labial *vs.* alveolar articulations (Figs 3 and 4 above). Comparisons between the mean frequencies using *t*-tests found no significant differences ($p > 0.05$) in the F2 frequencies of the final nasals in *seem to vs. been doing*. While we should be cautious about drawing inferences from an absence of difference, the data are sufficient in number to suggest strongly that /m/ in *seem to* assimilates to a high degree to the alveolar place of the following /t/. Figure 6 illustrates how F2 frequency of /m/ in *seem to* has a very similar distribution to that of F2 of /n/ in the control condition (before /t/ or /d/). This indicates a high rate of /m/-to-[n] assimilation in this word-pair.

In cases of assimilation like *seem to*, we might expect a bimodal distribution in the F2 patterns, with, for example, a large peak for the unassimilated tokens, and a distinct, smaller peak for assimilated tokens. However, we tend to observe rather broad, single-peaked distributions; this could be because the samples we have examined are from a diverse range of speakers. To test whether such evidence of assimilation might also be found in the speech of an individual and are not an accidental emergent property of

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

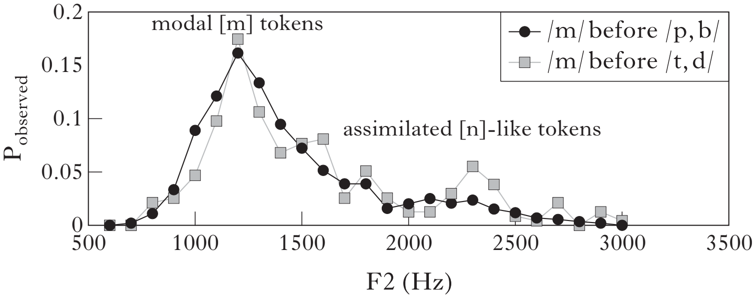


Figure 7

Histograms of F2 frequency of /m/ in all 1670 tokens produced by speaker 'Fred'. There were 235 tokens of /m/ before coronal consonants.

combining data from many speakers, Fig. 7 plots variation in the F2 frequency of all /m/ tokens within the speech of one selected speaker from the BNC, 'Fred', a 78-year-old man from the Northwest Midlands. This subject was chosen for in-depth analysis because he is heard to say *see*[n] *to* for *seem to* several times, even when repeating himself. In this figure, circles plot the F2 frequency distribution of 1435 unambiguously bilabial tokens, that is, /m/ tokens occurring prevocalically, utterance-finally or before labial consonants. In this control condition, there is a well-defined modal peak in the 1200–1300 Hz region. Squares indicate the F2 frequency distribution of 235 tokens of /m/ before coronal consonants. Again, a modal peak occurs in the 1200–1300 Hz region, showing that the /m/ tokens in this environment are also typically bilabial. However, additional smaller peaks are also evident at higher frequencies, due to a smaller proportion of /m/ tokens that have assimilated to their coronal context. Consistent with this interpretation, this speaker's tokens of *seem to* with an audible [n]-like consonant typically had a second mode of F2 frequencies between 1808 Hz and 2368 Hz. Nasals in the two contexts depicted have different mean F2s (/m/ control: 1473 Hz; /m/ before coronals: 1593 Hz). We compared the two distributions using a Kolmogorov-Smirnov test and found that they were significantly different from each other ($D = 0.238$, $p < 0.0005$). We also find statistical support for bimodality in the possible assimilation context: Hartigans' dip test for unimodality confirms that Fred's nasals before coronals are *not* unimodal ($D = 0.0565$, $p < 0.05$), while those in control contexts are unimodal ($D = 0.0067$, $p \sim 0.99$).

3.5.2 *Variation in velar nasals.* In the second planned comparison we examined whether the final /ŋ/ of *coming* varies according to (and in the direction of) a following /b/ or /d/. We compared the F2 frequency of the /ŋ/ in *coming back* with *coming down*. The nasal in *come back* is a control, as this is a context in which only bilabial variants are expected. As explained above, there are insufficient data even in this large dataset to make a

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

further planned comparison with a word-pair containing unassimilated /ŋ/. The means and standard deviations of the F2 measurements are given in Table VII.

	sex	tokens	F2	SD
come back	female	189	1579	277
/m/	male	185	1216	183
coming back	female	60	1730	235
/ŋ/ ~ [m]	male	65	1371	246
coming down	female	52	1858	245
/ŋ/ ~ [n]	male	47	1414	293

Table VII

Means and standard deviations (in Hz) of F2 frequencies of nasals.

Differences in mean F2 frequencies for *come back vs. coming back*, *come back vs. coming down* and *coming back vs. coming down* were compared using *t*-tests, with a Bonferroni correction for multiple comparisons. One-tailed significant differences in F2 frequency are summarised in Table VIII.

	mean F2 differences					
	coming back /ŋ/			*coming down* /ŋ/		
	sex	δ	p	sex	δ	p
come back	female	151	<0.0005	female	279	$<3\times10^{-10}$
/m/	male	155	<0.00001	male	198	<0.000001
coming back	—			female	128	<0.05
/ŋ/				male	43	*n.s.*

Table VIII

Size and significance of mean F2 differences (δ, in Hz)
for male and female speakers.

The differences between the F2 frequency distributions of /ŋ/ in these three word-pairs can also be seen in Fig. 8: in general, for both male and female speakers, the F2 of /ŋ/ before /b/ (*coming back*) is lower than before /d/ in (*coming down*), while the F2 of /m/ in *come back* (the control), as expected, is lower still. This is consistent with the pattern observed for word-final canonical /m/ *vs.* /n/ above (Figs 3 and 4), and indicates that a proportion of the tokens in each test context is assimilated

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

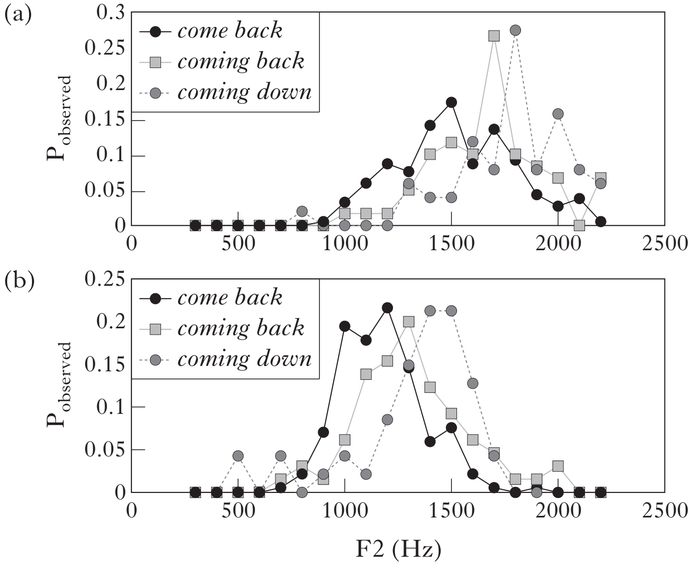


Figure 8

F2 variation in /m/ of *come back* and /ŋ/ of *coming back* and *coming down*. (a) Female speakers; (b) male speakers.

to the following consonant. The differences between mean F2 of the word-final nasals was significant in all pairwise comparisons, with the exception of *coming back vs. coming down* in the data of the male speakers. Taken together, these comparisons indicate that /ŋ/ does assimilate to some degree: along the F2 frequency scale there is a gradient of variation from canonical [m] in *come back* to variation between [ŋ] and [m] in *coming back* and to variation between [ŋ] and [n] in *coming down*.

As mentioned above, it is conceivable that assimilation in *coming* is coronal assimilation of the /-ɪn/ allomorph. Therefore, we also examined 89 tokens of *something but* and 92 tokens of *nothing but*, since these words are not affected by *-ing* allomorphy. In a listening test supported by visual examination of spectrograms, 6/89 (6.75%) of tokens of *something but* and 6/92 (6.5%) of tokens of *nothing but* were found to have unambiguous [m] at the end of *thing*. Since there is anecdotal evidence that some speakers have lexical alveolar nasals in these words too,[8] we identified all remaining tokens of *something* and *nothing* produced by the speakers of these twelve assimilated tokens. Figure 9 presents a selection of tokens of *something* and *nothing* produced by three of these speakers in both assimilation and non-assimilation contexts. The velar nasal evident in non-assimilation contexts shows that, for these speakers at least, the assimilation is from a velar to a bilabial place of articulation.

[8] We are grateful to one of the reviewers for challenging us with this observation.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

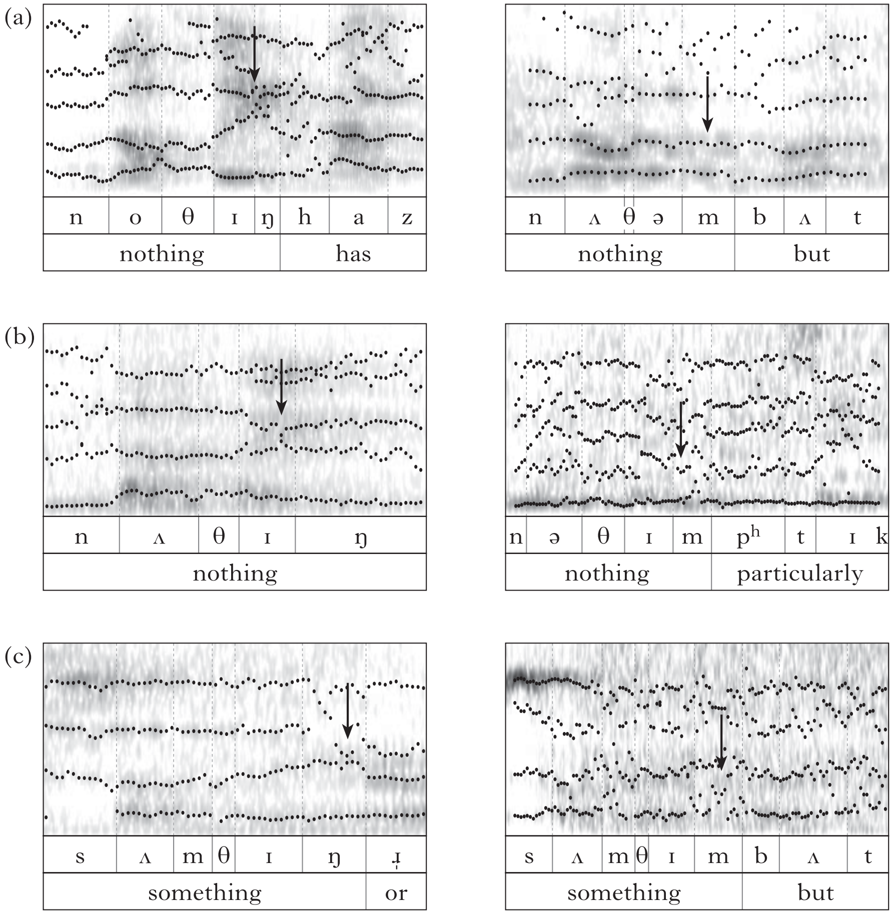


Figure 9

Unassimilated (left) and assimilated (right) forms of *nothing* and *something* in the speech of three male speakers: (a) speaker PS0S4; (b) speaker PS3KY; (c) speaker PS0LU. Arrows indicate the F2 of word-final nasals; tracks superimposed on the spectrograms show the lowest four formant frequencies, as estimated by Praat.

4 Accounting for variation in the place of articulation of nasals

'Coronal underspecification' (Avery & Rice 1989) offers an unambiguous prediction about word-final labials and velars: their place of articulation should not assimilate to that of following consonants. We have tested this prediction against over 15,000 tokens of relevant word-pairs from a corpus of natural English speech, and have found strong evidence that word-final bilabial and velar nasals do sometimes assimilate to following consonants.[9] For instance, in *coming back*, the mean F2 frequency of /ŋ/ is lower than in *coming down*, but not as low as the F2 frequency of /m/ in *come back*; in *something but* and *nothing but*, the F2 of /ŋ/ is similar to that of an [m]; and in *seem to*, the F2 of /m/ is as high as the F2 of /n/ in *been doing*, despite the fact that such an assimilation could lead to lexical confusion with *seen to* (and contrary to the theory that such distinctive contrasts should restrict phonetic variation; e.g. Lavoie 2002). Such examples conclusively demonstrate that word-final /m/ and /ŋ/ in English regularly assimilate to the place of articulation of the following consonant; nasal place assimilation is not restricted to coronals.

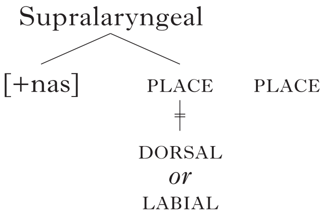
These counterexamples to the predictions of ‘coronal underspecification’ mean that phonological theory needs to be revised. In this section, we first consider how a rule- or constraint-based analysis of assimilation could be extended to include non-coronal assimilation. We then present an alternative probabilistic model, in which alveolars, bilabials and velars can *all* assimilate, but with different ranges of variation.

4.1 Rules *vs.* exceptions to coronal underspecification

As we have demonstrated, the constraint in (3) is contradicted by the fact that it is possible for /m/ and /ŋ/ to assimilate to following consonants. However, our data pose a more fundamental problem for underspecification in a standard rule- or constraint-based phonological framework: assimilation of non-coronal nasals to following coronals, as in *seem to* → *see*[n] *to*, cannot be expressed in the notation of coronal underspecification. For ‘assimilation by backward spreading’ to work in this case, the coronal consonant with which the second word begins must have some place of articulation content to spread backwards and thereby overwrite the word-final labial place feature(s). But if coronals have unspecified place, there will be no feature such as CORONAL to spread backwards onto the preceding nasal.

A possible way out of this difficulty is to establish a second – exceptional – rule to allow for *deletion* of word-final DORSAL or LABIAL place before a following coronal, as in (5).

(5)



{m, ŋ} → n / __ {t, d}

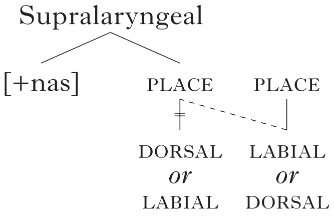
Having such an additional ‘exception’ rule has certain attractions. First, it could allow us to retain (2) and the generalisation that it captures, namely

[9] Assimilation of /m/ to [n] has also been observed in the Kiel Corpus of German by Zimmerer *et al*. (2009).

that assimilation is common in coronals. Second, it allows us to treat coronal and non-coronal assimilation as quantitatively different. For example, if we wish to associate probabilities or weights to the rules, as in the Variable Rule formalism of Cedergren & Sankoff (1974), there is no difficulty: (2) could be given a much higher weight or probability than (5). However, (5) still suffers from a severe problem, endemic to underspecification theory: here, 'empty PLACE' means 'specifically, invariably coronal', whereas in (2), empty PLACE means 'coronal, but may be filled by some other place specification', i.e. not *invariably* coronal. The problem, at root, is that underspecification theory overworks the use of empty nodes (Broe 1993: 203–206): 'empty' can mean 'the unmarked value of some contrast' – the emptiness is significant, as in (5) – or it can mean 'unspecified and thus susceptible to change by spreading', as in (2).[10]

A further weakness is that (5) *only* applies to following coronal contexts. To permit labials or velars to assimilate to following velars or labials respectively, additional rules complementary to (5) are needed, such as (6).

(6)

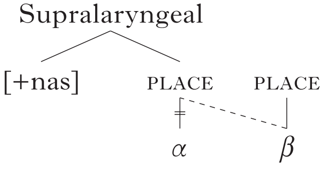


ŋ → m / __ {p, b}
m → ŋ / __ {k, g}

Now we have at least three rules: a main or default rule, (2), and two exceptions, (5) and (6). The addition of exceptions to deal with the new facts we have presented complicates the analysis.

(2), (5) and (6) miss the most important generalisation revealed in our data: in English, a word-final nasal with *any* place of articulation (in its citation/isolation form) can assimilate to *any* place of articulation of a following obstruent, as in (7).

(7)



Though (2) is not *false*, and does capture the most common case, it is essentially incorrect – because it is insufficient – as a *general* rule of place assimilation in English. On the other hand, though (7) captures all the cases, it does not reflect the fact that coronal assimilation, bilabial

[10] In fact, Broe (1993: 206) identifies *four* distinct 'meanings' of empty nodes in underspecification theory: predictable values, default values, potential but undefined values and undefinable values.

assimilation and velar assimilation do not occur with equal frequency. It is time for a rethink.

4.2 Probabilistic underspecification using Gaussian mixtures

Figures 3 and 4 above plot variation within a population, the set of Audio BNC speakers who happen to have uttered relevant word-pairs in our dataset. While the continuum of (acoustic) variation in the population is self-evident, we have no evidence that the *articulation* of /m/, /n/ or /ŋ/ in the speech of any individual varies continuously (apart from the usual variation due to coarticulation, which is small in comparison with the wide variations seen in our histograms). However, we have seen evidence of a bimodal distribution in the F2 data from one individual (Fig. 7). Furthermore, though variation in place of articulation between /n/ and /ŋ/ (i.e. via [n̠], [ɲ], etc.) is potentially continuous, the choice between the coronal and dorsal articulators is discrete. Moreover, articulatory continua between /ŋ/ and /m/ or between /n/ and /m/ are not physiologically conceivable. Thus, in producing, say, word-final /m/, a speaker alternates between distinct variants, which determine whether to make the oral closure using the lips, the tongue tip or the tongue dorsum, the choice being conditioned by the following context. The continuous but bimodal distribution of *seem to* and *see*[n] *to* shown in Fig. 6 above suggests that a probabilistic analysis is called for, for modelling assimilation within individual speakers as well as in a population sample.

Keating (1990) proposes an approach to modelling phonetic variation in which each target (on each articulatory dimension) is represented with a broader or narrower range, or window, of permitted variation. A narrow window presents a more specific, constrained articulatory target, while a broad window models underspecified, more variable articulations (Fig. 10a). Where the window of variation is very wide, quite large deviations from the typical pronunciation are possible. A model of this type supports constraints on variability such as canonically velar plosives admitting a very wide range of coarticulatory variation before following vowels (wide window or distribution = unconstrained, highly underspecified), whereas bilabial plosives have a much narrower range of variation (narrow window or distribution = more constrained, more tightly specified). This idea was implemented in a probabilistic form and used to model real articulatory data by Blackburn & Young (2000), as in Fig. 10b. Their probabilistic extension to Keating's model encodes the idea of a more probable central tendency, with permitted, but less likely, variants that may be rather different from the typical pronunciation. Blackburn & Young model the phonetics of each place of articulation specification using a Gaussian probability density function which is broader for 'underspecified' place and narrower for more tightly specified places. In Fig. 10b these Gaussian functions are drawn on the vertical axis because the figure incorporates a time dimension; in the figures we present elsewhere

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

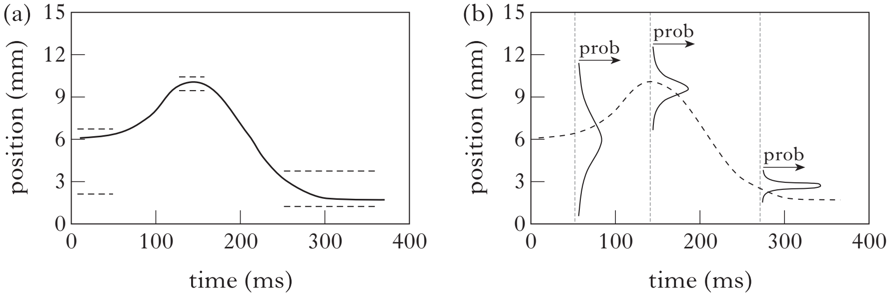


Figure 10

(a) Simulated articulator trajectory (solid line), using the window model of coarticulation (from Blackburn & Young 2000, after Keating 1990). The trajectory is constrained to pass through the 'windows' indicated by the dashed horizontal lines. (b) Simulated articulator trajectory (dashed line) using a probabilistic coarticulation model (from Blackburn & Young 2000). The midpoints of successive phonemes are indicated by the dotted vertical lines, and associated with each midpoint is a probability distribution, defining the probability that the articulator will take particular positions at the midpoints.

in this paper, the Gaussians relate to a single interval, the nasal consonant, and are therefore drawn horizontally.[11]

Although Blackburn & Young's version of Keating's model was proposed to account for coarticulation, it can be extended to model variation due to assimilation, provided that it is adapted to include potentially bimodal distributions in articulatory or acoustic variation. Rather than plotting the variation in articulatory positions, we use an acoustic measure, F2 frequency, as a proxy for place of articulation, just as in the distributions in Figs 3 and 4. Thus the Gaussian F2 distributions for bilabial, velar and alveolar place given in Figs 3b and 4b are probabilistically underspecified acoustic representations of the canonical place of articulation distinctions between /m/, /n/ and /ŋ/. Where assimilation leads to a categorical change in the articulator used, the data will contain some mixture of assimilated and unassimilated nasals. For example, the nasal in *seem to* is canonically bilabial, but a certain proportion of instances may be alveolar, making the distribution wider and possibly bimodal, whereas the nasal in *come back* is specifically bilabial, modelled with a narrower (more constrained, more tightly specified) window. We model such mixtures using weighted combinations of simple Gaussians: Gaussian

[11] Using a continuous statistical distribution to model acoustic phonetic variation within a category is far from novel, and is normal practice in automatic speech-recognition technology. In speech-perception research, logistic functions are often used to model perceptual variation within and between phonological categories. In a sociolinguistic context, Clopper (2014) uses probability density functions of F1 measurements on vowels to model the acoustic structure of vowel categories influenced by more than one dialect. See also note 12.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

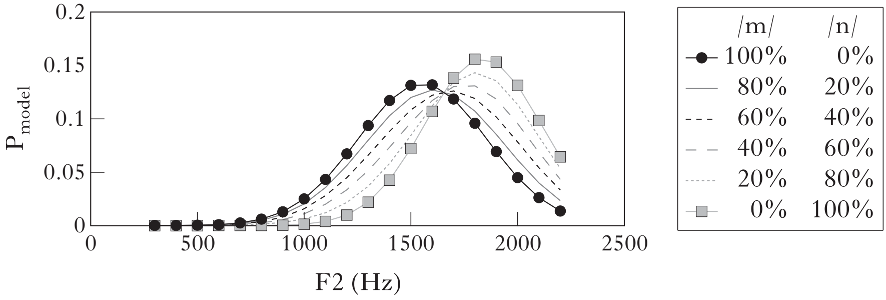


Figure 11

Gaussian models of /m/ and /n/, together with Gaussian mixture models (weighted sums of the Gaussian models of /m/ and /n/).

mixture models.[12] This is illustrated in Fig. 11, which shows the Gaussian models of /m/ and /n/ (female speakers), as in Fig. 3. Gaussian mixtures, created from weighted sums of the simple Gaussian models of /m/ and /n/ in varying proportion, are also plotted. It can be seen that the 80% /m/ + 20% /n/ mixture is quite close to the pure 100% /m/ Gaussian, the 20% /m/ + 80% /n/ mixture is close to the pure 100% /n/ Gaussian, and the other mixtures are, of course, in between. In order to find the appropriate proportion of *x*% /m/ + *y*% /n/ (or /ŋ/, as the case may be) for a given set of data, we fit a function as in (8) to the observed distribution, finding the values a_1 and a_2 for which the difference between the model and the observed data is minimised.

(8) $\mathrm{F2}_{\mathrm{observed}} = a_1\ \mathrm{probdf}(f, \mu_1, \sigma_1) + a_2\ \mathrm{probdf}(f, \mu_2, \sigma_2)$

The means and standard deviations μ_1, σ_1, μ_2 and σ_2 are obtained from the means and standard deviations of unassimilated /m, n, ŋ/ in Figs 3 and 4. μ_1 and σ_1 are the mean and standard deviation of the lexical nasal, and μ_2 and σ_2 are the mean and standard deviation of the nasal with the place of articulation of the following obstruent. *f* is the F2 frequency parameter.

Figure 12 illustrates how this approach models the combination of unassimilated and assimilated /m/ tokens in the nasal portion of the word-pair *seem to*, whose distribution is shown in Fig. 6 above. Fig. 12a plots simple Gaussian models for unassimilated /m/ and /n/, and a Gaussian mixture model for /m/ + /n/, i.e. a mixture of the unassimilated [m] and assimilated [n] variants of /m/ in *seem to*, based on the data from female speakers; Fig. 12b plots the corresponding models based on data from male speakers.

In like manner, Fig. 13 illustrates how this approach models the combination of unassimilated and assimilated /ŋ/ tokens in the nasal portion

[12] In a similar fashion, Goldrick *et al.* (2011) use mixtures of gamma distributions to model the range of variants in voiced *vs*. voiceless regions of the VOT continuum.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

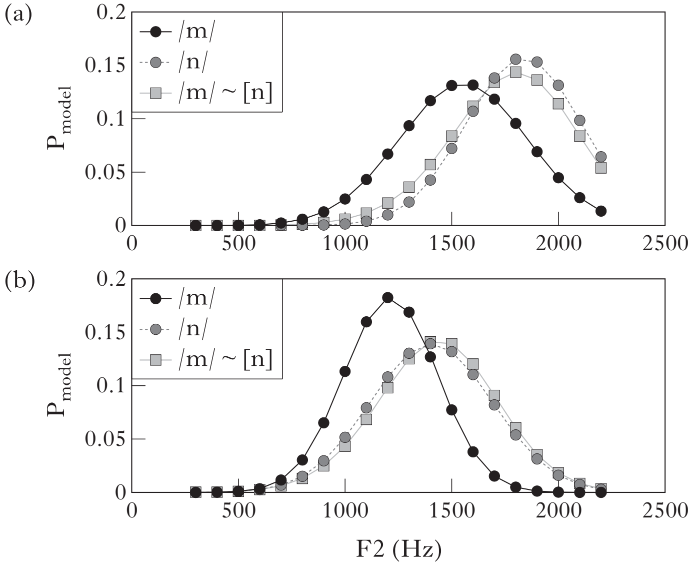


Figure 12

Modelling a combination of unassimilated and assimilated /m/ tokens, using Gaussian mixtures: (a) female speakers; (b) male speakers. Gaussian models of /m/ and /n/ are given for reference. Squares are Gaussian mixtures of /m/ and /n/ (female speakers: $a_1 = 1739$, $a_2 = 7453$, $\mu_m = 1554$ Hz, $\sigma_m = 302$ Hz, $\mu_n = 1838$ Hz, $\sigma_n = 271$ Hz; male speakers: $a_1 = -3705$, $a_2 = 18{,}183$, $\mu_m = 1214$ Hz, $\sigma_m = 218$ Hz, $\mu_n = 1404$ Hz, $\sigma_n = 287$ Hz).

of the phrases *coming back* and *coming down* in the speech of male and female speakers seen in Fig. 8. It shows Gaussian mixture models for /ŋ/ + /m/, i.e. a mixture of the unassimilated [ŋ] and assimilated [m] variants of /ŋ/ in *coming back*, and /ŋ/ + /n/, i.e. a mixture of unassimilated [ŋ] and assimilated [n] variants of /ŋ/ in *coming down*. The distributions of unassimilated and potentially assimilated tokens in these two figures illustrate how Blackburn & Young's probabilistic model of varying degrees of articulatory specification can be extended to model the assimilation of word-final nasals in different contexts: the phonological 'rule' is that a Gaussian mixture of the F2 distribution of lexical nasals and the F2 distribution of nasals with the place of articulation of the following consonant accurately models the variation observed in each assimilatory context, as expressed in (8) above. We discuss further in the following section why we view this as a phonological process.

4.3 Gaussian mixtures as phonological models of assimilation

The difference between (phonetic) coarticulation and (phonological) assimilation can sometimes be unclear. For example, the dentalisation of final coronal nasals before dental fricatives, e.g. [ɪn̪ðə] *in the*, is similar to

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

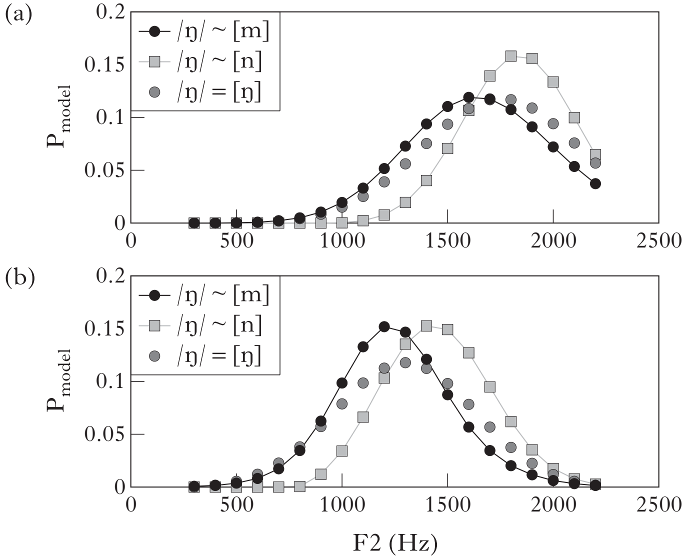


Figure 13

Modelling a combination of unassimilated and assimilated /ŋ/ tokens in *coming back vs. coming down*; (a) female speakers; (b) male speakers. Gaussian mixture models for /ŋ/, of the form a_1 probdf(f,μ_1,σ_1) + a_2 probdf(f,μ_2,σ_2), using the means and standard deviations of /ŋ, m, n/. Black circles are Gaussian mixtures of /ŋ/ and /m/ (female speakers: $a_1 = 3990$, $a_2 = 2979$, $\mu_ŋ = 1751$ Hz, $\sigma_ŋ = 371$ Hz, $\mu_m = 1554$ Hz, $\sigma_m = 302$ Hz; male speakers: $a_1 = 3028$, $a_2 = 3701$, $\mu_ŋ = 1299$ Hz, $\sigma_ŋ = 336$ Hz, $\mu_m = 1214$ Hz, $\sigma_m = 218$ Hz). Grey squares represent a Gaussian mixture of /ŋ/ and /n/ (female speakers: $a_1 = -359$, $a_2 = 5695$, $\mu_n = 1838$ Hz, $\sigma_n = 271$ Hz; male speakers: $a_1 = -2916$, $a_2 = 7791$, $\mu_n = 1404$ Hz, $\sigma_n = 287$ Hz). Grey circles represent simple (unmixed) Gaussian models of /ŋ/ for reference.

assimilation, but is often regarded as coarticulation because (a) there is no separate phoneme /n̪/ in English, and (b) it is variation within the single place category CORONAL. Similarly, the coarticulatory differences between the /k/ in *keep, cart* and *cool* are phonetic variations within the place category DORSAL. Assimilation of /m/ to [n], /m/ to [ŋ], /ŋ/ to [n] or /ŋ/ to [m] is of a quite different character: not variation within a place category, but involving alternation in the choice of articulator, as argued at the beginning of §4.2.[13] For example, Fig. 14 is a still image of United States President Barack Obama midway through saying *I'm* in *I'm gonna*

[13] It is possible for speakers to articulate a bilabial closure simultaneously with a coronal or dorsal one, such that the anterior articulation masks the posterior one. However, making a distinction between masked and unmasked tokens requires articulatory methodologies, and is thus not possible with an acoustic-phonetic corpus. Further video evidence of the type given in Fig. 14 would help to settle whether m → {ŋ, n} is gestural overlap or delabialisation, as in Fig. 14. The same issue, of course, applies to distinguishing between masked and unmasked tokens of /n/, which is generally accepted as undergoing assimilation.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

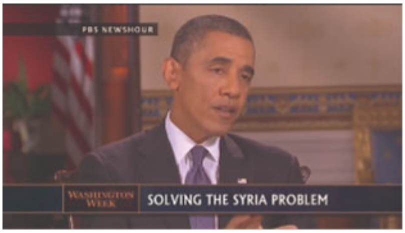


Figure 14
Image of President Obama midway through saying *I'm* in *I'm gonna convince*. (Source: PBS Newshour/YouTube; http://www.youtube.com/watch?v = s4OwubYrL2c#t=9m24s). An AVI file of the *I'm gonna* clip is available in the online version of the paper, and at http://www.phon.ox.ac.uk/jcoleman/Obama_I_m_gonna.avi.

convince; this is one of four video frames in which he articulates the medial [ŋ] of [ʌŋənə]. In all of those frames, it is quite evident that his lips never close, as they would for an [m]; there is no bilabial closure in this token – he uses a distinctly different articulator, indicating a categorical assimilatory switch, not a gradient process of coarticulation.

The pronunciation of /m/ as [n] or [ŋ] and of /ŋ/ as [m] or [n] thus clearly constitutes assimilation. Moreover, we have amassed a range of empirical evidence demonstrating that such assimilation, though a little rarer than coronal assimilation, is systematic in our large corpus of natural speech data. It must therefore be a function of the phonological knowledge of our speakers, just as the possible articulation of /n/ as [m] or [ŋ] is generally accepted to be. As we have seen (§4.1), although it is possible to model it by introducing supplementary abstract rules (or constraints), this becomes excessively complicated and ends up masking the important generalisation that all word-final nasals may assimilate.

We have shown how Keating's model of coarticulation can be extended and adapted to model assimilation using Gaussian mixtures. The simple Gaussian models from which the mixtures are composed are abstractions over the phonetic data, just as a category such as [+voice] is an abstraction over a range of VOT values. While each individual F2 datum for /ŋ/ is, clearly, a phonetic event, we regard a distribution such as $\mathrm{probdf}(f,\mu_{\text{ŋ}},\sigma_{\text{ŋ}})$ as a single phonological primitive (in this case, a specification of place of articulation). The Gaussian mixtures constructed from such objects encode the probability of choice between two or more such phonological possibilities in a given context.

6 Conclusion

The size of the Audio BNC has allowed us to show, contrary to numerous earlier statements, that word-final /m/ and /ŋ/ can and sometimes do assimilate in English to the place of articulation of following word-initial obstruents, a fact that is inconsistent with the phonological theory of coronal underspecification. Non-coronal assimilation is found in the speech of a large number of speakers. It is detectable auditorily, and visible in spectrograms. Moreover, F2 frequency varies according to the place of articulation of following consonants in the ways we expect it to, and these patterns are statistically robust. This is as systematic for /m/ and /ŋ/ as it is for /n/. The strength of phonetic evidence for non-coronal assimilation indicates that it is real, yet its relative rarity calls for a new kind of analytical framework, in which the different frequencies of assimilation are explicitly encoded probabilistically. We have shown how histograms of observed F2 distributions of canonical, unassimilated 'control' forms can be modelled using Gaussian probability density functions, and how the F2 distributions of nasals in assimilation context can be modelled using Gaussian mixtures of those canonical forms. This new, probabilistic approach extends Keating's (1990) 'phonetic underspecification' model to cover cases of phonological assimilation, i.e. the optional selection of a distinct place of articulation. This model is an abstraction over the data, just as phonological rules are. It is an underspecification model because it does not specify which place of articulation will be selected in any given instance, instead allowing for a range of contextually conditioned variants. It has two major advantages over the rule/constraint-based model: firstly, the same model can be applied to word-final /m, n, ŋ/, without the need for supplementary rules for exceptions. Secondly, because it is probabilistic, it captures not simply the fact that assimilation may occur in a given context (as do rules (2) and (7)), but also the likelihood of its occurring for a given nasal in a given context. It is therefore more descriptively adequate.

REFERENCES

Avery, Peter & Keren Rice (1989). Segment structure and coronal underspecification. *Phonology* **6**. 179–200.

Baayen, R. Harald, D. J. Davidson & D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**. 390–412.

Baghai-Ravary, Ladan, Greg Kochanski & John Coleman (2011). Data-driven approaches to objective evaluation of phoneme alignment systems. In Zygmunt Vetulani (ed.) *Human language technology: challenges for computer science*. Berlin: Springer. 1–11.

Barry, Martin C. (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics* **4**. 1–16.

Bates, D. M. & Martin Maechler (2009). Package 'lme4': linear mixed-effects models using S4 classes. https://cran.r-project.org/web/packages/lme4/lme4.pdf.

Bird, Steven (1992). Finite-state phonology in HPSG. In *COLING-92: Proceedings of the 15th International Conference on Computational Linguistics*. Vol. 1. 74–80.

Blackburn, C. Simon & Steve Young (2000). A self-learning predictive model of articulator movements during speech production. *JASA* **107**. 1659–1670.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

BNC Consortium (2007). *BNC XML edition*. Available at http://ota.ox.ac.uk/desc/2554.

Boersma, Paul & David Weenink (2012). *Praat: doing phonetics by computer* (version 5.3.35). http://www.praat.org.

Broe, Michael (1993). *Specification Theory: the treatment of redundancy in generative phonology*. PhD dissertation, University of Edinburgh. Available (August 2016) at http://www.phon.ox.ac.uk/files/pdfs/Broe1993.pdf.

Cedergren, Henrietta J. & David Sankoff (1974). Variable rules: performance as a statistical reflection of competence. *Lg* **50**. 333–355.

Clopper, Cynthia G. (2014). Sound change in the individual: effects of exposure on cross-dialect speech processing. *Laboratory Phonology* **5**. 69–90.

Coleman, John, Ladan Baghai-Ravary, John Pybus & Sergio Grau (2012). *Audio BNC: the audio edition of the Spoken British National Corpus*. Phonetics Laboratory, University of Oxford. http://www.phon.ox.ac.uk/AudioBNC.

Crowdy, Steve (1993). Spoken corpus design. *Literary and Linguistic Computing* **8**. 259–265.

Crowdy, Steve (1994). Spoken corpus transcription. *Literary and Linguistic Computing* **9**. 25–28.

Crowdy, Steve (1995). The BNC spoken corpus. In Geoffrey Leech, Greg Myers & Jenny Thomas (eds.) *Spoken English on computer: transcription: mark-up and application*. London: Longman. 224–234.

Cruttenden, Alan (2014). *Gimson's pronunciation of English*. 8th edn. London: Routledge.

Dilley, Laura C. & Mark A. Pitt (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *JASA* **122**. 2340–2353.

Fujimura, Osamu (1962). Analysis of nasal consonants. *JASA* **34**. 1865–1875.

Goldrick, Matthew, H. Ross Baker, Amanda Murphy & Melissa Baese-Berk (2011). Interaction and representational integration: evidence from speech errors. *Cognition* **121**. 58–72.

Harris, John (1994). *English sound structure*. Oxford: Blackwell.

Keating, Patricia A. (1990). The window model of coarticulation: articulatory evidence. In John Kingston & Mary E. Beckman (eds.) *Papers in laboratory phonology I: between the grammar and physics of speech*. Cambridge: Cambridge University Press. 451–470.

Kiparsky, Paul (1985). Some consequences of Lexical Phonology. *Phonology Yearbook* **2**. 85–138.

Kreidler, Charles W. (1989). *The pronunciation of English*. Oxford: Blackwell.

Kurowski, Kathleen & Sheila E. Blumstein (1987). Acoustic properties for place of articulation in nasal consonants. *JASA* **81**. 1917–1927.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen (2013). lmerTest: tests in linear mixed effect models. *R package* (version 2.0). https://cran.r-project.org/web/packages/lmerTest.

Labov, William (1989). The child as linguistic historian. *Language Variation and Change* **1**. 85–97.

Lavoie, Lisa M. (2002). Subphonemic and suballophonic consonant variation: the role of the phoneme inventory. *ZAS Papers in Linguistics* **28**. 39–54.

Lodge, Ken (2009). *A critical introduction to phonetics*. London & New York: Continuum.

McMahon, April (2002). *An introduction to English phonology*. Edinburgh: Edinburgh University Press.

Mandelbrot, Benoit (1961). On the theory of word frequencies and on related Markovian models of discourse. In Roman Jakobson (ed.) *Structure of language*

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press

and its mathematical aspects. Providence, RI: American Mathematical Society. 190–219.

Manuel, Sharon Y. (1995). Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *JPh* **23**. 453–476.

Miller, George A. & Noam Chomsky (1963). Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.) *Handbook of mathematical psychology*. Vol. 2. New York: Wiley. 419–491.

Ogden, Richard (1999). A declarative account of strong and weak auxiliaries in English. *Phonology* **16**. 55–92.

Olive, Joseph P., Alice Greenwood & John Coleman (1993). *Acoustics of American English speech: a dynamic approach*. New York: Springer.

Potter, Ralph K., George A. Kopp & Harriet Green Kopp (1966). *Visible speech*. 2nd edn. New York: Dover.

Renwick, Margaret E. L., Ladan Baghai-Ravary, Rosalind A. M. Temple & John Coleman (2013). Assimilation of word-final nasals to following word-initial place of articulation in United Kingdom English. *Proceedings of Meetings on Acoustics* **19**. 060257. http://dx.doi.org/10.1121/1.4800279.

Repp, Bruno H. & Katyanee Svastikula (1988). Perception of the [m]–[n] distinction in VC syllables. *JASA* **83**. 237–247.

Roca, Iggy & Wyn Johnson (1999). *A course in phonology*. Oxford & Malden, Mass.: Blackwell.

Shockey, Linda (2003). *Sound patterns of spoken English*. Malden, Mass. & Oxford: Blackwell.

Stevens, Kenneth N. (1998). *Acoustic phonetics*. Cambridge, Mass: MIT Press.

Trudgill, Peter (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Young, Steve, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying (Andrew) Liu, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev & Phil Woodland (2009). *The HTK Book* (*for HTK Version 3.4*). Available (August 2016) at http://staffhome.ecm.uwa.edu.au/~00014742/research/speech/local/htk/htkbook.pdf.

Yuan, Jiahong & Mark Liberman (2008). Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08*. 5687–5690. Software release at https://www.ling.upenn.edu/phonetics/old_website_2015/p2fa/.

Yuan, Jiahong & Mark Liberman (2011). Automatic detection of 'g-dropping' in American English using forced alignment. In *Proceedings of 2011 IEEE Workshop on Automatic Speech Recognition and Understanding*. 490–493. Available (August 2016) at http://www.ling.upenn.edu/~jiahong/publications/cn1.pdf.

Zimmerer, Frank, Henning Reetz & Aditi Lahiri (2009). Place assimilation across words in running speech: corpus analysis and perception. *JASA* **125**. 2307–2322.

Zipf, George Kingsley (1935). *The psycho-biology of language: an introduction to dynamic philology*. Boston: Houghton Mifflin.

https://doi.org/10.1017/S0952675716000208 Published online by Cambridge University Press