



Research Article

Sandra Kotzor*, Allison Wetterlin, Adam Charles Roberts,
Henning Reetz and Aditi Lahiri

Bengali nasal vowels: lexical representation and listener perception

<https://doi.org/10.1515/phon-2022-2017>

Published online May 27, 2022

Abstract: This paper focuses on the question of the representation of nasality as well as speakers' awareness and perceptual use of phonetic nasalisation by examining surface nasalisation in two types of vowels in Bengali: underlying nasal vowels (C \tilde{V} C) and nasalised vowels before a nasal consonant (CVN). A series of three cross-modal forced-choice experiments was used to investigate the hypothesis that only unpredictable nasalisation is stored and that this sparse representation governs how listeners interpret vowel nasality. Visual full-word targets were preceded by auditory primes consisting of CV segments of CVC words with nasal vowels ([tʃã] for [tʃãd] 'moon'), oral vowels ([tʃa] for [tʃal] 'unboiled rice') or nasalised oral vowels ([tʃã(n)] for [tʃãn] 'bath') and reaction times and errors were measured. Some targets fully matched the prime while some matched surface or underlying representation only. Faster reaction times and fewer errors were observed after C \tilde{V} C primes compared to both CVC and CVN primes. Furthermore, any surface nasality was most frequently matched to a C \tilde{V} C target unless no such target was available. Both reaction times and error data indicate that nasal vowels are specified for nasality leading to faster recognition compared to underspecified oral vowels, which cannot be perfectly matched with incoming signals.

Keywords: hearer perception; lexical representation; nasal vowels; phonology

***Corresponding author: Sandra Kotzor**, Language and Brain Laboratory, Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, UK,
E-mail: sandra.kotzor@ling-phil.ox.ac.uk

Allison Wetterlin, University of Agder, Kristiansand, Norway

Adam Charles Roberts, Future Resilient Systems Singapore-ETH Centre, Singapore, Singapore.
<https://orcid.org/0000-0001-6497-0717>

Henning Reetz, Institut für Phonetik, Goethe Universität Frankfurt, Frankfurt, Germany

Aditi Lahiri, Language and Brain Laboratory, Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, UK

1 Introduction

1.1 Background

The phonetics and phonology of nasal vowels have been studied extensively and, while an accurate objective measurement of nasality is difficult to obtain, several theoretical approaches have been put forward as to the phonetic characteristics of nasality, the cues in the speech signal marking nasality used by listeners, and their lexical representations (see Krämer 2017: 408ff for a review). Theories of how listeners abstract information from the speech signal and recognise speech sounds and words can be said to differ in two fundamental ways. First, they differ in which information is stored for any given word in the mental lexicon. The theoretical proposals range from models that assume listeners store very rich episodic information for every word every time it is encountered (Johnson 1997; Pierrehumbert 2002) to models that are more parsimonious and store only information that is contrastive and where each word or morpheme only has one underlying representation (e.g. Norris et al. 2000; Lahiri and Reetz 2002, 2010). Second, especially in the latter group, theories differ in terms of which information is extracted from the speech signal to access the mental representations, for example articulatory events (e.g. Liberman and Mattingly 1985; Fowler 1986), phonemes (e.g. Marslen-Wilson 1987, 1989; Norris 1994; Norris and McQueen 2008), or features (e.g. Gaskell and Marslen-Wilson 1997; Lahiri and Reetz 2002, 2010). One of the focal debates amongst the latter models, and one of the questions investigated in this paper, centres on the issue of the specification of nasality in the lexicon and whether the feature [NASAL] is privative or equipollent.

Chomsky and Halle (1968: 316) treat [NASAL] as a binary feature, proposing nasal [+nasal] and nonnasal [–nasal]. More complex representations have been proposed in recent years where only marked, prominent or active features are represented (e.g. Clements 2001), suggesting that languages differ in their representation. For instance, Cohn (1993) suggested that, since French contrasts nasal and oral vowels, these ought to be represented as [+nasal] and [–nasal] respectively, while this is irrelevant for English vowels. Clements, however, would distinguish between lexical representation and phonologically active features, such that a lexical representation may lack a specification to begin with, but a feature may become active later in the phonology. Similarly, Trigo (1993: 393) proposes, in an investigation of Lusitanian Portuguese, that the feature NASAL is equipollent ([+nasal] vs. [–nasal]) in languages where it ‘crucially distinguishes among two or more elements’ (e.g. in Bengali) while it is privative ([NASAL] vs. no

value) in languages where this is not the case (e.g. in English).¹ In addition, in equipollent systems, there have also been conflicting proposals regarding the representation of contextually nasalised vowels, with some suggesting that those vowels are represented as [+nasal] (cf. Ohala and Ohala 1995 for English CVN words) or [–nasal] (cf. Chomsky and Halle 1968).

From a theoretical standpoint, an equipollent or binary system which refers to nasal sounds as [+nasal] and consequently assumes that oral vowels and consonants are specified as [–nasal] would presuppose that all items which are specified for either [+nasal] or [–nasal] form a coherent class with certain properties common to them all. It would also follow that both features could be active in phonological processes such as spreading. The assumptions for a privative or monovalent system are rather different: for instance, the opposite of the feature [NASAL], inherent to all oral vowels and consonants, does not form a coherent set since it is not specified (sometimes indicated by empty brackets []; Lahiri 2018). There is no one feature which groups together all oral vowels and oral consonants and thus there is no featural information which is extracted from the signal to match with a mental representation of oral sounds or to be active in phonological processes. One example of a theoretical lexical access account which proposes a monovalent system and operates on a feature level rather than using phonemes (unlike, for example the Cohort model; cf. Marslen-Wilson 1987, 1989) is the Featureally Underspecified Lexicon (FUL; Lahiri and Reetz 2002, 2010).

In addition to the question regarding the representation of nasality, this paper also addresses whether listeners distinguish between underlying and contextual nasality, and if so, which cues they use to discriminate between the two. The present study is not primarily concerned with the question of the extent to which the contextually nasalised vowel in a CVN word (e.g. English *ban*) differs phonetically from an underlying nasal vowel, but with the question of what native listeners of a language with both underlyingly nasal and contextually nasalised vowels do when they perceive **any** nasality in a vowel. If one assumes that there is a phonetic difference between contextual and underlying nasals (e.g. Cohn 1993; Beddor et al. 2013), hypotheses concerning the interpretation of nasality from the signal would differ for the two types of vowels. If phonetic cues are unambiguous as to the source of the nasality, then listeners would surely make use of these cues and thus would be

¹ Scholars differ in their representation of nasality. McCarthy (1988), for instance, has [NASAL] attached directly to the root node, as is also true for Halle et al. (2000). However, no special claim is made about the status of nasality as a privative feature (or not). Dresher (2009) maintains that feature specification is built on contrastive hierarchies and uses binary values but is not necessarily committed to them. Hyman (1973) makes a very clear typology for languages with and without nasal contrasts but also ends up assuming a binary distinction. In sum, specific language analyses have required the choice of monovalent or privative features.

able to accurately identify the source of the nasality in the signal (i.e., an underlying nasal vs. a contextually nasalised segment). If, on the other hand, representation governs perception and recognition, any nasality in the signal would in the first instance lead listeners to an underlying, fully represented nasal vowel.

There has been strong evidence that listeners are affected by top-down effects when processing the acoustic signal. Both likelihood and lexical status of a perceived signal (e.g. Ganong 1980) as well as the phonological system of a listener's native language(s) have been found to guide perception and can lead to listeners seemingly disregarding certain acoustic cues (Jongman et al. 1992). Ganong's seminal paper (1980) showed that categorisation of stimuli with ambiguous VOT could depend on the existence of a word. Thus, the same VOT duration would be categorised as [t] if the decision was between *tack*/**dack* but as [d] if the choice was between **tesk*/*desk*. Jongman et al.'s experiment also revealed that identical vowel duration led to significant variation in choosing between words and nonwords based on their phonological representation. They appended different word-initial consonants to a continuum from [at] to [a:t] to create two pairs of real words in Dutch: /zat/ ~ /za:d/ 'drunk' ~ 'seed' and /stad/ ~ /sta:t/ 'city' ~ 'state'. Since Dutch voiced stops are devoiced word finally, the surface pairs would be [zat] ~ [za:t] and [stat] ~ [sta:t]. The results showed that the vowel categorisation between long and short depended on the underlying phonological representation of the voicing of the final consonants:

The phoneme boundary for /zat/ ~ /za:d/ occurred at a significantly shorter vowel length than /stad/ ~ /sta:t/. Thus, the results showed that the categorisation of ambiguous vowel length was affected by the underlying voicing of the following word final consonant.

The fact that phonological representations guide perception has been demonstrated in studies across different languages as well as dialects (i.e. Pallier et al. 1997; Pallier et al. 2001; Scharinger and Lahiri 2010), which show that phonetically identical contrasts are interpreted differently by speakers of different languages/dialects due to differences in the phonological features that are represented in their lexicon. Scharinger and Lahiri (2010), for example, show how the different specification of short front vowels for [LOW] in two dialects of English (American and New Zealand) affected listeners' processing in a semantic priming study.

In the present study we are not, however, dealing with variability across dialects or languages but with the inherently variable phonetic property of nasality. Thus, our central question is whether listeners ascribe surface nasality to regressive assimilation from a following nasal consonant or whether an interpretation as an underlying nasal vowel would be equally, or indeed more, likely in languages such as Bengali, which have both contextually nasalised and underlyingly nasal vowels. Thus, in addition to the question of how nasality is represented, we are

also concerned with the question whether listeners' choices are indeed governed by this representational information. While there has been some previous work in this area (cf. Section 1.3 below), we are using different methods to elicit quicker decisions from listeners without any context in order to shed light on automatic access to stored representations.

1.2 Acoustic properties of vowel nasality

1.2.1 Measures of nasality

There is a substantial literature on the production, perception, and acoustics of nasal and contextually nasalised vowels and the phonetic differences which may present listeners with cues to distinguish between oral and nasal vowels. Although measuring acoustic cues of nasality is not the central focus of our study, we briefly provide an overview of the relevant acoustic properties that have been suggested in the literature to gauge nasality. The complex structure of the nasal tract (Bjuggren and Fant 1964; Dang et al. 1994; Pruthi et al. 2007) with asymmetric passages and additional side-cavities leads to complex acoustic effects (Dang and Honda 1996; Lindqvist-Gauffin and Sundberg 1976). The main effects for nasality found in the speech signal are flattening and widening of the oral first formant F1 (Fant 1960; House and Stevens 1956; Stevens 1998: 193), additional nasal poles and zeros which are above F1 for high vowels and below F1 for low vowels (Hattori et al. 1958; Hawkins and Stevens 1985; Maeda 1982, 1983); see also Carignan 2018, for an extended summary) and increased low-frequency energy (House and Stevens 1956; but see Vampola et al. 2020) alongside other effects (see e.g. Styler 2017). Although the nasal tract is essentially a rigid structure, even the location of the poles and zeros of the nasal tract, as manifested in the acoustic signal, depends on the amount of opening of the velopharyngeal port (Pruthi 2007: 76) (in contradiction to Stevens' description (1998: 306)). Usual measures of nasality are the bandwidth of F1 (B1), amplitudes of the nasal poles (P0, P1) and the relation of these amplitudes to the amplitude of the first formant (A1) (Chen 1996, 1997) or relations of these amplitudes to those of the first two harmonics (H0, H1) (Huffman 1990). These measures are generally taken from the amplitude of harmonics closest to F1 (A1) and the frequencies of the nasal poles (FP0, FP1), respectively. Since the formant frequencies of vowels are close to those of the nasal poles, Chen (1996: 129) proposed an adjustment formula for the related amplitudes to account for differences in the effect of vowel types on nasal peak amplitude. Nevertheless, the closeness of F0, F1, FP0 and FP1 (and F2 for back vowels) can make it impossible to determine formant frequencies independently,

especially for high-pitched voices where the wide spacing of the harmonics does not provide the granularity to separate these frequencies (see e.g. Chen 1996: 130; Berger 2007: 16). Styler (2017) proposed a procedure for measuring several parameters automatically, rather than hand-locating the respective resonances manually.

The efficiency of the above measures, however, varies considerably across speakers. For example, Berger (2007: 50) found overall best discrimination results between nasal and oral vowels for the difference A1–H1, followed by A1–P0 and B1 for Bengali, English and Spanish. Note that Bengali vowels were underlyingly nasal while English and Spanish had nasalised vowels in a nasal context. But A1–P0 was the second worst out of the six parameters he investigated for the female Bengali speaker in his study. Pruthi (2007) investigated 37 different measures on three databases (TIMIT,² StoryDB,³ ICSI⁴) and found that overall in an ANOVA comparing oral and nasalised vowels, B1 yielded the highest F-ratios, followed by the A1–H1 measure and one A1–P0 measure.^{5,6} It is worth noting that in all such investigations the widening of the bandwidth of F1 (B1) was one of the best scoring parameters related to nasality, a fact already observed by House and Stevens (1956). According to the Source-Filter Theory (Fant 1960), we can assume that the quality of the vocal tract filter (Quality or Q-factor: formant centre frequency divided by its bandwidth) is reasonably constant across a wide frequency range, as long as the velum is raised. Lowering the velum will change this quality by introducing the resonating and shunting cavities of the nasal tract, which leads to a decrease in the Q-factor. We therefore suggest Q1 (F1/B1) as our preferred measure of nasality, which can be compared across vowels independent of vowel-specific variation of F1. The lower the Q1 measure of the vowel, the more nasality is present.

1.2.2 Phonetic differences between contextual and underlying nasals

Many studies have tackled the question of phonetic differences between contextually nasalised and underlying nasal vowels. Krämer (2017) reviews the literature

² Six hours of single sentences read by 438 male and 192 female speakers of American English.

³ Seven vowels in nasalised and non-nasalised contexts in 56 real words, spoken by one American English speaker (whose vocal tract had been measured by means of MRI) in four conditions and with four repetitions (= 896 words).

⁴ Part of the switchboard corpus of spontaneous telephone recordings in the USA (Byrne et al. 1998; Godfrey et al. 1992) by an unknown number of male and female speakers; 5086 sentences.

⁵ He tested five different methods and selected one with the highest F-ratio comparing oral and nasalised vowels (Pruthi 2007: 100).

⁶ The ranking differed between the three databases.

on the phonetics of English nasalised vowels and the comparisons with languages with both underlying nasal vowels (e.g. French (Chen 1997)) and contextualised nasal vowels (e.g. Spanish (Solé 1992, 1995) or Thai (Beddor 2007)) and concludes that most production studies do not find less robust nasalisation for English contextually nasal vowels compared to underlying nasal vowels in other languages (Cohn (1993) is the notable exception here). However, much of this literature deals with English and a comparison of the contextually nasalised vowels in English to underlying nasals in other languages, rather than a comparison of how listeners in a language which has both contextually nasal and underlying nasal vowels interpret nasality (for a review of both production and perception studies on English nasalised vowels see Krämer (2017)).

Although it is difficult to find consistent spectral properties that can be attributed to nasality in all vowels (vowel-independent properties (Beddor 1993: 173)), we know that listeners perceive and use nasality in vowels in languages both with and without an oral-nasal vowel contrast (Butcher 1976; Wright 1986) to assign vowels to two distinct categories (Beddor and Strange 1982). Furthermore, for more than eight decades it has been consistently shown that listeners use vowel nasalisation as a cue for an upcoming nasal consonant (cf. Malécot 1960). However, it has also been demonstrated that, phonetic cues aside, a listener's phonology plays an important role in how nasality is used. Lahiri and Marslen-Wilson's (1991) study of Bengali and English listeners showed that nasalisation in vowels is only used as a cue for an upcoming nasal consonant in speakers of a language where nasality is predictable (English) and not for speakers where vowel nasality is contrastive (Bengali).

1.3 Lexical representation of nasality

Lexical access models predominantly focus on access routes to the stored information and often do not provide much detail about the representation of features, such as nasality, or phonemes but only present a brief overview of the phonological information contained in the mental representations (see McQueen 2005: 268 for an overview of proposed representations in common speech perception models). Models such as TRACE (Elman and McClelland 1986; McClelland and Elman 1986) or Cohort (Marslen-Wilson and Welsh 1978; Marslen-Wilson 1993) and the Distributed Cohort Model (DCM, Gaskell and Marslen-Wilson 1997) as well as Stevens' (2002) Acoustic Landmarks Model assume a featural representation either at the pre-lexical stage or as underlying representations. An account which assumes featural underspecification would be compatible with models proposing featural representations, while models such as Shortlist A and B (Norris 1994;

Norris and McQueen 2008) and exemplar-based models (Johnson 1997; Pierrehumbert 2002), which propose phonemic representations, would result in different processing predictions. The present study is primarily concerned with how contrastive features are stored and we provide predictions based on underspecified representations, phonemic representations, and exemplar-based representations in detail below.

As far as the question of the representation of vowel nasality is concerned, earlier research on Bengali shows support for $[_{NASAL}]$ as a privative feature, claiming that predictable nasality is not represented (Lahiri and Marslen-Wilson 1991, 1992; henceforth L&M-W). Contextually nasalised vowels in CVN sequences would thus be represented without nasality, e.g. [bān] ‘flood’ from underlying /ban/. This results in ambiguous surface forms in languages like Bengali where there are also underlying nasal vowels. These cases can only be disambiguated by the following consonant. In the case of an oral vowel in the sensory input (e.g. [baḍ] ‘omit’), an underlyingly nasal vowel would no longer be a candidate since this combination results in a no-mismatch between the sensory input (underspecified) and the representation ($[_{NASAL}]$). Nasal or contextually nasalised vowels in cases like Bengali (e.g. [bāḍ^h] ‘dam’, [bān] ‘flood’), however, should not mismatch with the underlyingly oral vowel since oral vowels have no featural specification for nasality.

L&M-W (1991) used a gating task (‘t Hart and Cohen 1964; Grosjean 1980) with incremental presentation of the CV segment of CVC words where listeners were asked to respond with a full word which forms a continuation of the phonetic CV sequence they were perceiving. Their stimuli were sets of triplets in Bengali (CVC [baḍ], CVN [bān], and CVC [bāḍ^h]) as well as doublets where no word with a nasal vowel exists in the same consonantal environment (CVC [lob^h] ‘greed’ and CVN [lōm] ‘body hair’).

L&M-W found that in the triplet condition both stimuli with contextualised nasal vowels and underlying nasal vowels, i.e. CVN and CVC stimuli, respectively elicited a large percentage of CVC responses at the offset of the vowel.⁷ This shows that listeners take nasalisation as a cue that they have heard a CVC rather than a CVN word. CVN responses for both categories were very low (below 8%) but this changes rapidly as soon as consonantal information becomes available since this disambiguates the surface nasality. The best match for the nasality in the auditory signal was a word with a vowel underlyingly specified for nasality, i.e. a CVC word. CVN words are still an option since there is no mismatch between the signal and the underspecified underlying representation of the oral vowel. This data shows support for underspecification since a surface hypothesis would have predicted an

⁷ Mean percentage of responses up to gate 0 (offset of vowel) is used for all percentages given here.

equal distribution of CVC and CVN responses here as both vowels would have been specified for nasality.

CVC stimuli resulted in a large proportion of CVC responses (>80%) even at very early gates.⁸ CVC responses were almost non-existent (0.7%) but there are a relatively large number of CVN responses (13.4%) considering the lack of nasality in the signal (L&M-W 1991). This phenomenon is difficult to explain with a surface representation account since, with nasality present in both representations, there should be no difference between CVN and CVC responses to CVC stimuli. This pattern is more easily explained by proposing an underlyingly oral vowel in CVN words. CVN responses to CVC stimuli are thus not only permitted but, based on the underlying representation alone, should be as likely as CVC responses since they have the same representation (L&M-W 1991). The fact that there are more CVC responses than CVN responses corresponds to the proportion of CVC and CVN patterns in the language, with CVC sequences being considerably more frequent than CVNs.

Data from L&M-W's doublet sets exhibited similar patterns for the CVC stimuli, and the CVN stimuli result in a different pattern where listeners produced 65% CVC responses, 16% CVN responses and 17% CVC responses even though no CVC item, in this consonantal frame, is available in the lexicon of the language. In these cases, participants often responded with CVC words from Hindi, another language they were familiar with. The increase in CVC and CVN responses compared to the triplet set shows that the information in the signal matches both representations and there is no other competitor containing nasality (unlike in the triplet case). This makes the distribution of responses more similar to the CVC stimuli responses above, which are in part determined by pattern frequency in the language.

In a subsequent study using Hindi and English, Ohala and Ohala (1995) attempted to replicate L&M-W's findings with methodological modifications since they had several criticisms of the original study, including the fact that in L&M-W's study no statistical analysis was possible since participants' responses were not constrained, and that previously heard versions at earlier gates could have influenced participants' responses. Ohala & Ohala restricted the participants' responses and, in order to prevent previous stimuli affecting subsequent responses, presented only one gated version of each stimulus per participant (with the exception of the most severely gated stimulus, which was followed by the full word after four trials). While they found results similar to those of L&M-W (1991) in their Hindi triplet condition as well as with CVN stimuli in the Hindi doublet condition, thereby lending support to the underspecification hypothesis, they do not see

⁸ Responses do not add up to 100% as not all responses were scoreable (see Lahiri and Marslen-Wilson 1991: 269).

either gating study as providing convincing evidence for underspecification of nasality but state that ‘at best [these findings] are compatible with both the UR [underlying representation] and SR [surface representation] hypotheses’ (Ohala and Ohala 1995: 57).

Both studies (Lahiri and Marslen-Wilson 1991, 1992; Ohala and Ohala 1995) thus showed that in a gating paradigm listeners displayed an overwhelming preference for contrastively nasal vowels in response to any nasality in the signal (CVN or C \check{V} C), to the extent that, in cases where there was a lexical gap (e.g. in Lahiri & Marslen-Wilson’s doublet conditions), listeners provided words with nasal vowels in response to CVN fragments from another language they were familiar with (Hindi).

1.4 Present study and predictions

To address the central questions of this study – whether nasality is represented as an equipollent or a privative feature, and how listeners use **phonetic and phonological** nasality in processing – we conducted a series of experiments using a cross-modal (audio-visual) form priming task with a forced-choice response paradigm (Cooper et al. 2002). All stimuli (described in greater detail in Section 2.1) were monosyllabic Bengali words belonging to one of three types: C \check{V} C, CVN and CVC. The CV/ \check{V} segments of these words were used as primes, with the full words as targets.

Before setting out the predictions resulting from different types of representational hypotheses, we briefly discuss the distributional patterns of nasal and non-nasal vowels in Bengali.

The Bengali vowel system consists of seven oral vowel phonemes and seven nasal counterparts (cf. Chatterji 1926; Ferguson and Chowdhury 1960). All oral vowels can also become nasalised before the three nasal consonants /m, n, ŋ/. Thus, oral vowels are, evidently, the most frequent as they can (theoretically) occur with any of the language’s 30 oral consonants. While nasal vowels can occur preceding all consonants apart from the nasals, they are most frequent before coronal consonants. As contextually nasalised vowels can only occur in certain environments (preceding one of the three nasals in the language), these are the most restricted and therefore least likely. However, these frequency assessments, based on the combinatorial possibilities afforded by the phonological inventory of the language, only tell part of the story, since it also matters how frequent the words are in which these patterns occur. Many of the words with the CVN pattern are highly frequent even though the combinatorial possibilities are the most severely restricted.

Depending on the representational specifications (i.e. privative vs. equipolent) as well as the precise specification of vowels in CVN contexts and the surface cues in the acoustic signal, several hypotheses are possible. We outline below one set of surface (phonetic) predictions (cf. Figure 1) based on the acoustic cues (assuming no difference between contextually nasalised and underlyingly nasal vowels) and one based on a monovalent system where CVN is unspecified for nasality (cf. Figure 2). In a surface account, where listeners are able to directly match primes and targets based on purely phonetic cues, we would expect faster latencies only for identity pairs since this approach would result in only one possible match per prime (cf. Figure 1).

In a monovalent system, however, where nasal vowels (C \tilde{V} C) are specified as [NASAL] while oral vowels (CVC) and vowels nasalised due to a following nasal (CVN) are not specified for nasality, shorter response latencies as well as lower error rates should be observed after C \tilde{V} C primes (compared to CVC and CVN primes) since their representation better matches the surface form. In Figure 2, we lay out in greater detail the predictions made by an underspecification account assuming a monovalent feature [NASAL]. As far as this approach is concerned, the only genuine match is that between a surface nasal vowel (from either C \tilde{V} C or CVN)

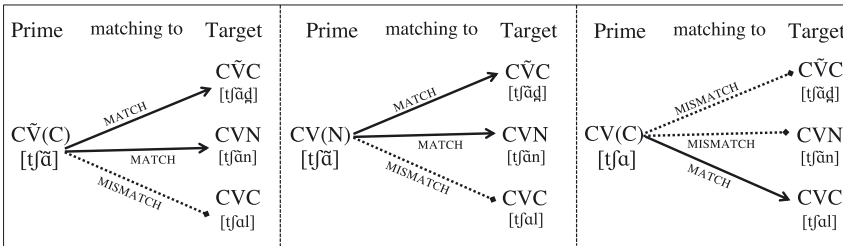


Figure 1: Surface phonetic predictions for C \tilde{V} (C), CV(N) and CV(C) primes.

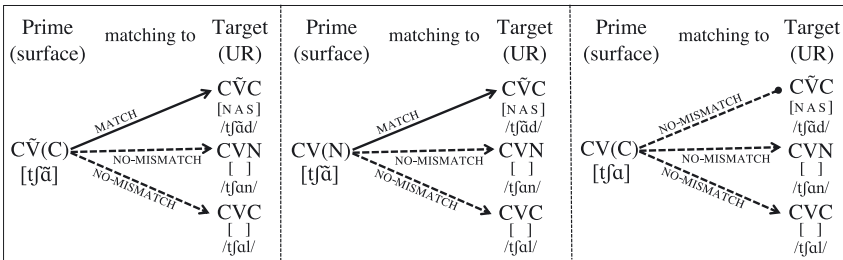


Figure 2: Phonological predictions for match and no-mismatch relationships for C \tilde{V} (C), CV(N) and CV(C) primes in an underspecification account.

and an underlyingly specified nasal vowel /C \tilde{V} C/ (cf. Lahiri and Reetz, 2002, 2010). All other conditions should result in a no-mismatch scenario between the input and the underlying representation since no nasal or oral feature is specified in either CVC or CVN words.

This results in predictions of only two real match conditions across the possible prime-target combinations. We expect an identity match between the underlying nasal vowel prime (C \tilde{V} C) and its identity target and another match between a surface nasal vowel from a CVN prime and the specified nasal vowel target (C \tilde{V} C). Faster response latencies and lower error rates are predicted for match conditions than for no-mismatch conditions (Lahiri and Reetz 2002). In terms of the error data, however, it is unclear whether the distinction between match and no-mismatch will surface clearly. Surface nasalisation may well play a more crucial role in the error data than in the reaction times, which are a more gradient measure. Since nasality is highly salient, this may lead to more accurate matching in cases where the prime contains a nasal or contextually nasalised vowel and a word with either option (C \tilde{V} C or CVN) is present as a target.

In approaches which are fundamentally guided by frequency and experience (such as an exemplar model), it is never precisely articulated how one particular exemplar is chosen out of a number of possible options. Frequency will contribute to this selection as will, presumably, surface cues, such as the degree of nasalisation of a vowel. Since the degree of nasalisation can vary greatly in both underlying nasals and contextual nasals and exact matches are not required in an exemplar model, there must be a gradient approach and nasalisation in the signal could lead to either an underlying nasal or a contextually nasalised vowel, depending on other contributing factors such as, for instance, frequency. It is thus difficult to provide clear-cut predictions for such an exemplar model: these predictions would fall somewhere in between those made by a fine-grained phonetic account and an underspecification approach, but they would be clearly distinct from both. Phonetic accounts would only lead to an identity match, and an underspecification account makes the clear prediction that, given any nasality in the signal, the first choice would be an underlying nasal. as this is the only option specified for nasality.

In terms of frequency, two different aspects need to be considered. Firstly, if nasal vowels are heard more frequently in a C \tilde{V} C rather than a CVN frame (as there are only three nasal consonants in Bengali which would trigger regressive nasalisation), listeners should be faster to match nasal vowel primes (be it from a C \tilde{V} C or CVN prime) to C \tilde{V} C targets. However, since the oral vowel only occurs in CVC contexts, matching these to CVC targets should be fastest overall since this is the most frequent pattern. If the number of neighbours is a determining factor here, a faster response might be expected for C \tilde{V} C prime-target combinations as these

have fewer neighbours. CVN words have even fewer neighbours and should thus result in even faster reaction times than CVC words, with CVC words being reacted to slowest.

There is no published corpus for Bengali to allow us to compute neighbourhood density and vowel frequency statistics. Basic lexical frequency counts were taken from a Bengali corpus of three million words (Dash and Chaudhuri 2001) and mean frequencies for each condition can be found in Table 1. However, these can be misleading as the lexical distribution is not the same. The number of words with nasalised vowels in the context of nasal consonants /m n ŋ/ is, by definition, smaller than that of words with oral vowels which can be followed by any of the 30 oral consonants. However, in contrast to (for instance) French, and other languages where nasality has been studied, Bengali is, to our knowledge, one of the few languages which has a full set of contrastive nasal and oral vowels (seven each) (Ferguson and Chowdhury 1960: 32; Hajek 2013; Klaiman and Lahiri 2018: 423). Consequently, Bengali provides the best possible foundation for an investigation of the phonological representation of nasality.

Table 1: Prime-target examples of doublet and triplet sets by block (with NEITHER cases indicated in bold print) including frequency counts taken from a corpus of three million words (Dash and Chaudhuri 2001).

Prime	Triplets			
	CVC [tʃā(d)]	CVN [tʃā(n)]	CVC [tʃa(l)]	
Target (Block 1)	tʃān - tʃād tʃal - tʃān tʃād - tʃal	tʃān - tʃād tʃal - tʃān tʃād - tʃal	tʃān - tʃād tʃal - tʃān tʃād - tʃal	
Target (Block 2)	tʃād - tʃān tʃān - tʃal tʃal - tʃād	tʃād - tʃān tʃān - tʃal tʃal - tʃād	tʃād - tʃān tʃān - tʃal tʃal - tʃād	
Mean freq	235.08	108.23	246.23	
Prime	Doublets (NoCVC)		Doublets (NoCVN)	
	CVN [tʃī(n)]	CVC [tʃī(l)]	CVC [dʒ ^h ā(p)]	CVC [dʒ ^h a(l)]
Target (1)	tʃīn - tʃīl	tʃīl - tʃīn	dʒ ^h al - dʒ ^h āp	dʒ ^h āp - dʒ ^h al
Target (2)	tʃīl - tʃīn	tʃīn - tʃīl	dʒ ^h āp - dʒ ^h al	dʒ ^h al - dʒ ^h āp
Mean freq	506.17	193.07	33.21	300.21

2 Materials and design

To test the two sets of predictions laid out in Figures 1 and 2, we investigated the processing of nasality on vowels (phonemic as well as underlyingly nasal) with one priming and two word-recognition tasks. The material was the same in all tasks and is described here for the priming task first.

We compared the processing of the vowels with different competitor sets in a cross-modal (audio-visual) form priming task with a forced-choice response paradigm where participants had to choose one of two forms presented in Bengali script. We chose three sets of stimuli: two *doublet sets* where either the CVN or CVC pattern is not attested in the language and one *triplet set* where all three options exist. Thus, in the two doublet sets, there are either no CVC competitors for CVN words or no CVN competitors for CVC words. In the case where all three patterns exist in the language (i.e. the triplet set), there are either one matching target for the prime and one other competitor (a MATCH condition) or no matching item in the target pairs (a NEITHER condition).

2.1 Stimuli

2.1.1 Primes

CV fragments of 42 common monosyllabic CVC words of Bengali were used as primes and they were divided into three sets of near-minimal pairs: one set of 14 triplets and two sets of 14 doublets each (for examples see Table 1). The triplet set contained CVC strings in which all possible variations of vocalic nasality led to real words in the language: oral vowels (e.g. [tʃal] চাল ‘unboiled rice’), nasal vowels (e.g. [tʃɑ̃d̪] চাঁদ ‘moon’) and nasalised oral vowels due to a following nasal (e.g. [tʃɑ̃n] চান ‘bath’). In the first doublet set (NoCVC), words with the same consonant sequence containing a nasal vowel do not exist in Bengali and thus the two primes in this set contained oral vowels (e.g. [tʃil] তিল ‘sesame seeds’) and nasalised oral vowels due to a following nasal (e.g. [tʃin] তিন ‘three’). The second doublet set (NoCVN) contained only oral vowels (e.g. [dʒʰal] ঝাল ‘spicy hot’) and underlying nasal vowels (e.g. [dʒʰɑ̃p] বাঁপ ‘jump’), and in these cases the corresponding minimally different CV pairs with nasalised oral vowels are not attested in Bengali.

All primes were recorded in full and then truncated using PRAAT (Boersma and Weenink 2011) to include the complete duration of the vowel while ensuring no consonantal information is available. This was done based on the spectral properties; obstruents were cut at the beginning of the closure and for sonorants,

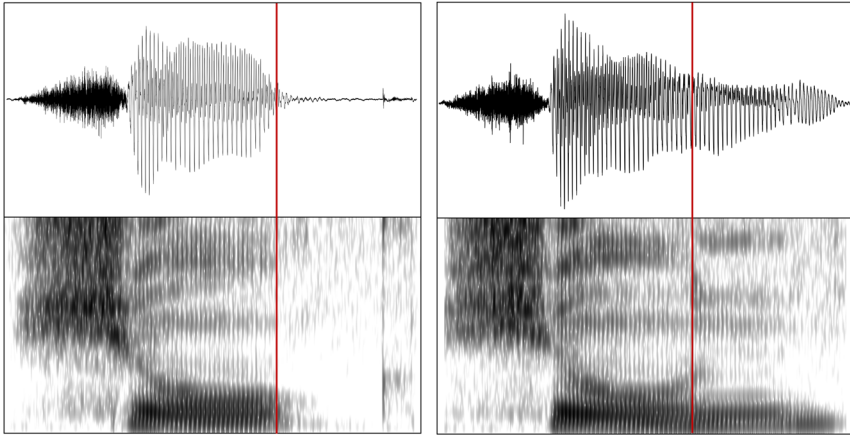


Figure 3: Example of full CVC word /jōk/ (left panel) and CVN word /jon/ (right panel). The lines mark the ends of the prime fragments /jō/ and /jo/ respectively.

decisions were based on close examination of the spectrogram as well as our own perception (cf. Figure 3 for an example of full words and truncated primes). There were no significant length differences of the individual vowels between the three conditions (oral, nasalised, underlyingly nasal).

2.1.2 Targets

Targets were the complete words matched to the fragment primes. In the doublet conditions both possible targets were presented, while in the triplet condition only two of the possible three targets were presented (cf. Table 1). Thus, in one third of the cases neither of the two targets matched the prime, resulting in two conditions for the triplet set: one where an identity match was possible (MATCH) and one where neither target was an identity match (NEITHER). Targets were presented in pairs in two blocks with words in alternate positions (see Section 2.2 below).

2.1.3 Stimulus recording

All primes were recorded by a female native speaker of Bengali in a sound-attenuated room with a Roland R-26 WAV recorder at a sampling rate of 44.1 kHz using a high-quality microphone (Shure SM27). The words were extracted and digitised, and the volume equalised using PRAAT (Boersma and Weenink 2011).

2.1.4 Acoustic measurements of nasality

As discussed in Section 1.2.1, we believe that Q1 is the most reliable vowel- and speaker-independent acoustic correlate of nasalisation. However, since other measures have been preferred in the literature, we therefore report the *A1–PO compensated* and *A1–P1 compensated* measures provided by Styler (2017) as well as the *Q1* measure. We computed these values for each vowel at five points throughout the vowel (at 20, 40, 60, 80 and 100% of the vowel duration) using a modified PRAAT script (Boersma and Weenink 2011) from Styler (2017).⁹

The measure *A1–PO compensated* grouped the nasal and oral vowels together as oral, while the nasalised vowels were calculated as being the ‘most nasal’, and *A1–P1 compensated* could not differentiate between the three different vowel types at all. Table 2 shows the results of these three measures at five positions in the vowel for all three different stimulus types (CVN, CVC and CVC) as well as the results of *t*-tests to compare all possible combinations within a measure. For the

Table 2: Average *Q1*, *A1PO compensated*, and *A1P1 compensated* measures at five timing points for all vowels in CVC, CVN, and CVC stimuli and results of *t*-tests (Tukey HSD) comparing those measures (* indicates significance).

		20%	40%	60%	80%	100%
CVC	Q1	3.42	2.86	2.59	2.71	3.07
CVN	Q1	3.02	2.51	1.76	1.90	2.66
CVC	Q1	5.98	4.30	4.13	4.47	4.52
CVN/CVC	<i>t</i>	$p = 0.9031$	$p = 0.8803$	$p = 0.3798$	$p = 0.3276$	$p = 0.8207$
CVN/CVC	<i>t</i>	$p = 0.0018^*$	$p = 0.0251^*$	$p = 0.0002^*$	$p < 0.0001^*$	$p = 0.0090^*$
CVC/CVC	<i>t</i>	$p = 0.0079^*$	$p = 0.0090^*$	$p = 0.0021^*$	$p = 0.0029^*$	$p = 0.0521$
CVC	a1p0 comp	0.34	1.24	0.26	0.10	-1.21
CVN	a1p0 comp	-0.15	-0.76	-1.34	-3.51	-2.97
CVC	a1p0 comp	4.13	1.40	0.61	0.72	0.58
CVN/CVC	<i>t</i>	$p = 0.9587$	$p = 0.1917$	$p = 0.3098$	$p = 0.0219^*$	$p = 0.3987$
CVN/CVC	<i>t</i>	$p = 0.0251^*$	$p = 0.1001$	$p = 0.1282$	$p = 0.0022^*$	$p = 0.0140^*$
CVC/CVC	<i>t</i>	$p = 0.0538$	$p = 0.9863$	$p = 0.9355$	$p = 0.8685$	$p = 0.3235$
CVC	a1p1 comp	15.85	17.85	18.49	18.82	17.34
CVN	a1p1 comp	15.80	17.78	20.45	19.11	23.63
CVC	a1p1 comp	19.37	17.29	18.60	19.31	21.78
CVN/CVC	<i>t</i>	$p = 0.9999$	$p = 0.9998$	$p = 0.8853$	$p = 0.9972$	$p = 0.1713$
CVN/CVC	<i>t</i>	$p = 0.5085$	$p = 0.9861$	$p = 0.8779$	$p = 0.9984$	$p = 0.8278$
CVC/CVC	<i>t</i>	$p = 0.5176$	$p = 0.9817$	$p = 0.9995$	$p = 0.9904$	$p = 0.3460$

⁹ https://github.com/stylerw/styler_praat_scripts/tree/master/nasality_automeasure (last accessed on 20th January 2022).

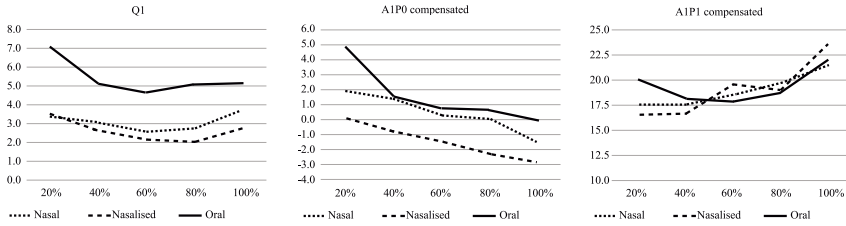


Figure 4: Average Q1, A1P0 compensated, and A1P1 compensated measures at five timing points for all vowels in $C\check{V}C$ (nasal), CVN (nasalised), and CVC (oral) stimuli.

same stimuli, the Q1 measure groups the nasalised and the underlying nasal vowels together, differentiating them from the oral vowels (Figure 4).

This data show that there are only small differences in the Q1 values between the vowels of $C\check{V}C$ and CVN stimuli at any of the timing points (not even at 20% or 40%, where CVN vowels may be expected to show less nasality) while the Q1 values differ at all timing points between oral (CVC) vowels and both underlyingly nasal ($C\check{V}C$) and contextualised nasal vowels (CVN). This demonstrates that there are no significant acoustic differences in the degree of nasalisation between $C\check{V}C$ and CVN stimuli, while both are acoustically distinct from CVC items. We further analysed the formant transitions of F1 and F2 as well as how frequently a target with a specific final consonant was selected for a given input fragment which determined that there were no additional consistent phonetic cues which could influence the main result. The results of the formant transitions show a considerable number of differences for F1 (as indicator of manner of articulation) between oral vowels and nasal vowel. For F2, there are only few differences, mostly for high vowels (which are few in the experiment). Thus, the coarticulation points to different manners of articulation, while differences in place of articulation are not evident in this formant analysis. The results of the analysis based on a targets' final consonant indicate that participants choose alternative words mostly on the basis of manner of articulation rather than based on place information. This contradicts the information available from the formant transitions. Detailed results for these analyses can be found in the Supplementary material to this article.

2.1.5 Participants

Fifty-nine female native speakers of Bengali (aged 18–23; mean average age 19.67) participated in the priming experiments. All were undergraduate students at Gokhale Memorial Girls' College (Kolkata) and had corrected-to-normal vision and no hearing impairments. The participants were compensated appropriately for their participation.

2.2 Procedure

In the priming experiment, participants were tested in groups of at most 16 in a quiet and darkened room. The auditory primes were played through individual on-ear headphones (SONY MDR110 LP) and visual targets were projected onto a screen in Bengali script with a data projector. Reactions and reaction times were captured using a custom-made multi-participant experimental setup (Reetz and Kleinmann 2003). In a standard forced-choice paradigm, participants were asked to indicate whether the CV fragment they heard belonged to the left or right target word on the screen. No information was given about the oral or nasal quality of the signals. Responses were made via individual custom-made two-button boxes with the left button corresponding to the left target and the right to the right target. All participants were right-handed and used their dominant hand to indicate if the prime matched the target displayed on the right. A ten-item practice task with items not used in the main experiment was conducted and re-run if necessary until it was clear participants were comfortable with the task.

Trials were separated into two blocks, which were presented to different participants. Every prime was presented once in the case of the doublets and three times in the triplet set in each of the two blocks, to ensure that each possible prime-target combination was only shown once per block (cf. Table 1). Each block consisted of 182 pseudo-randomised trials presented in a different order in each block (Table 1 shows all 13 combinations in each block for one set of the 14 items). The side of presentation (left vs. right) of the target words was randomised and counterbalanced across blocks (cf. Table 1). Participants heard one beep followed by a 200 ms silence before each auditory prime. Immediately at the end of the prime the visual targets (pairs of words) were displayed for 800 ms. After an additional 700 ms (i.e., 1500 ms after the end of the prime and the beginning of the visual display) the next trial started with its beep. After every 14 trials a sequence of three beeps was presented, which served as a short break. Participants were told they would hear the beginning of a word and then had to decide which of the words presented on the screen was the closest continuation for what they had heard. They were instructed to respond as quickly and accurately as possible.

2.3 Word comprehension tasks

Two word comprehension tasks were conducted after the priming experiment to test listeners' perception of the stimuli used in the that experiment. In both cases, listeners were presented with the same CV/√ fragments used in the priming experiment, presented in a completely randomised order. The auditory fragments

were played through headphones at a fixed pace. Participants were asked to respond in writing (pen and paper task) with the syllable they heard (Test 1; $n = 26$) or the complete monosyllabic word the fragment was taken from (Test 2; $n = 26$).¹⁰ The results of these tasks are presented in Section 3.1 below.

2.4 Analysis of priming experiment

The data from two participants had to be omitted due to a malfunctioning button box and one additional participant was removed from the analysis due to consistently fast and even negative reaction times suggesting they were reacting to the prime rather than the target. In addition, all reaction times ≤ 0 ms and outside ± 2 standard deviations were excluded from the following analyses (4.24% of data for triplet conditions and 3.96% for doublets). Errors (response accuracy) were analysed first separately, comparing the proportion of correct responses to a chance proportion of 50% using Pearson's Chi Squared tests. Errors were then analysed using logit generalised linear models, treating responses (error/correct) as a binomial distribution. Reaction times were analysed using linear mixed effects models.

3 Results

After the presentation of the results of the word recognition tasks, the results section discusses both RT and error data for the doublet and triplet conditions of the priming experiment separately. First the analyses of the triplet data will be presented, followed by the presentation of the doublet data, which investigates how listeners use nasality when there is a lexical gap in the language and one variant of the CVC sequence (either $C\check{V}C$ or CVN) is not available in the lexicon. Errors were then analysed using logit generalised linear models, treating responses (error/correct) as a binomial distribution, for the fixed effect of PRIME VOWEL ($CV(N)$ versus $C\check{V}$, $CV(N)$ versus CV , $C\check{V}$ versus CV), with random intercepts specified for PARTICIPANTS and TARGETS. Reaction times were analysed using linear mixed effects models, again for the fixed effect of PRIME VOWEL, with random intercepts specified for PARTICIPANTS and TARGETS.

¹⁰ Three participants did not take part in the word comprehension experiment.

3.1 Word recognition task results

As can be seen in Figure 5, CV syllable and CVC words result in the most accurate responses (91.4 and 75.3%), while CV(N) syllables seem to be most difficult for listeners to identify accurately from CV(N) fragments (2.6%). Even CVN words gave only 34.6% correct responses, which may well be modulated by the words available in the lexicon and their frequencies. This indicates that, in the syllable task, the nasality which is present in the signal of a CV(N) fragment is not unambiguous and is interpreted by listeners as either a CV or a C \tilde{V} syllable (49.4 and 48% responses).

For CV and C \tilde{V} syllables and words, listeners are more likely to choose the correct word rather than any other and while there are a significant number of C \tilde{V} syllable and C \tilde{V} C word responses (36.7 and 30.8%) to C \tilde{V} fragments, there are hardly any or only few CV(N) responses (0.3 and 16.2% respectively).

3.2 Priming triplet results

Since, as mentioned above, in one third of the trials there was no possible identity match of prime and target, the triplet data will be analysed in two separate conditions: MATCH and NEITHER. The MATCH condition includes all instances where an

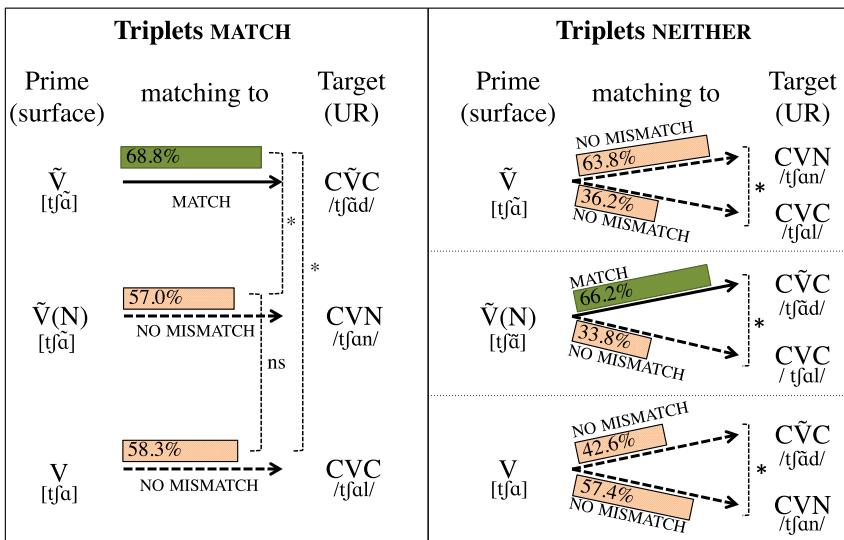


Figure 5: Predictions and error results for triplet match (% correct) and neither (% chosen) conditions.

exact match to the prime was available as a target (e.g. prime [tʃã] (from [tʃãḍ]); targets: [tʃãḍ] – [tʃal]) and here only correct trials are included in our analysis. The NEITHER condition contains the data from those trials where no identity match was available (e.g. prime [tʃã] (from [tʃãḍ]); targets: [tʃãḥ] – [tʃal]) and therefore, since there is no correct item, all trials are analysed.

In the MATCH condition, errors were analysed using a logit generalised linear model treating responses (error/correct) as a binomial distribution, for the fixed effect of PRIME VOWEL (CV(N), CṼ, CV), with random intercepts specified for PARTICIPANTS and TARGETS. Reaction times were analysed using a linear mixed model for the fixed effects of PRIME VOWEL and CONTEXT OF RESPONSE (nested under PRIME VOWEL, i.e., the target that did not match the prime; Prime ~ Context: CV ~ CṼC, CV ~ CVN, CṼ ~ CVC, CṼ ~ CVN, CV(N) ~ CVC, CV(N) ~ CṼC), with random intercepts specified for PARTICIPANTS and TARGETS.

In the NEITHER condition, response preferences were analysed using a logit generalised linear model treating responses (possible response A/possible response B) as a binomial distribution, for the fixed effect of PRIME VOWEL, with random intercepts specified for PARTICIPANTS and TARGETS. Reaction times were analysed using a linear mixed model for the fixed effects of PRIME VOWEL and CONTEXT OF RESPONSE, with random intercepts specified for PARTICIPANTS and TARGETS.

3.2.1 Reaction time results

In the MATCH condition, the linear mixed model shows an overall significant effect for PRIME VOWEL ($F(2, 260.5) = 17.93, p < 0.001$) as well as for the CONTEXT OF RESPONSE ($F(2, 171.2) = 5.09, p = 0.002$). In post-hoc Tukey's HSD tests between the three different prime vowels, we see significant differences between the response latencies to targets after CṼ versus CV(N) primes ($p = 0.001$) and CṼ versus CV primes ($p < 0.001$) with responses after CṼ primes being faster than those after CV(N) and CV primes. The difference between latencies after CV(N) and CV primes is not significant ($p = 0.065$). When the context of the response is taken into account (i.e. whether the other target option was a CVN, CVC or CṼC word), the only significant difference in the planned comparisons is found after CV(N) primes, where participants are significantly faster to make an identity match if the other target is a CṼC word ($t(171.2) = -2.88, p = 0.005$) while there is no such difference for CVC ($t(171.2) = -1.42, p = 0.157$) or CṼC ($t(171.3) = -0.41, p = 0.680$) identity matches (see Table 3 for details).

In the NEITHER condition, the linear mixed model showed an overall significant effect for PRIME VOWEL ($F(2, 83.07) = 4.63, p < 0.012$) and the effect of CONTEXT OF RESPONSE was significant ($F(3, 1968) = 14.33, p < 0.001$). A planned comparison between the two different possible responses for each prime vowel resulted in the following

Table 3: Triplet reaction time results (combined RT for correct responses and identity responses by context (i.e. the other available prime) in the MATCH condition and combined responses as well as RT per chosen target in the NEITHER condition).

		MATCH					
Prime		CVC		CVN		CVC	
Context of response		CVN	CVC	CVC	CVC	CVC	CVN
RT (ms)		555	556	569	597	608	586
(StdError)		(13.14)	(12.93)	(13.73)	(13.05)	(13.40)	(13.10)
RT (ms) overall		555		583		597	
(StdError)		(11.26)		(11.47)		(11.38)	
		NEITHER					
Prime		CVC		CVN		CVC	
chosen		CVC	CVN	CVC	CVC	CVC	CVN
RT (ms)		614	636	567	622	587	611
(StdError)		(14.27)	(13.35)	(13.32)	(14.45)	(13.95)	(13.48)
RT (ms) overall		625		594		599	
(StdError)		(12.96)		(12.98)		(12.91)	

pattern: If the prime vowel was CV(N), participants reacted significantly faster when they chose a CVC target compared to a CVN target ($t(1968) = -5.58, p < 0.001$). When the prime was a CV, participants chose a CVC target significantly faster than a CVN target ($t(1968) = 2.62, p = 0.009$), and similarly when the prime contained a CVC, the reactions to CVC were significantly faster than to the CVN targets ($t(1968) = 2.31, p = 0.021$). Both these cases are no-mismatch scenarios and, with no underlying phonological cues to guide responses, the results here may well be affected by frequency, as the CVN condition in the triplet set is overall less frequent than the CVC and CVC conditions. While all conditions show significant differences in listeners' preferences for a certain response, they show a significantly stronger preference (for the CVC target) when the prime is a CV(N) fragment (see Figure 5 for an overview). This can be seen by comparing the degrees of reaction time difference across conditions (i.e. difference of differences), we find that responses to CV(N) primes are significantly more different than those to CV primes ($t(2010) = -5.85, p < 0.001$) and CVC primes ($t(2010) = -5.61, p < 0.001$), while there is no difference between the CV and CVC primes ($t(2010) = -0.17, p = 0.866$).

3.2.2 Error analysis

Overall, in the *MATCH* condition, participants responded with the identity target 61.4% of the time and performed significantly better than chance in this task ($\chi^2(1) = 208.21, p < 0.001$). In analyses by *PRIME VOWEL*, participants' responses were significantly more accurate than chance in all three conditions: *CVC* ($\chi^2(1) = 189.89, p < 0.001$), *CVN* ($\chi^2(1) = 26.52, p < 0.001$) and *CVC* ($\chi^2(1) = 36.70, p < 0.001$).

In a logit generalised linear model (fixed effect: *PRIME VOWEL*), the data shows a significant effect for *PRIME VOWEL* ($\chi^2(2) = 48.06, p < 0.001$). In a planned comparison of the percentages of correct responses to targets we find a significant difference between responses following *CṼ* versus *CV(N)* primes ($\chi^2(1) = 40.16, p < 0.001$) as well as between those after *CṼ* versus *CV* primes ($\chi^2(1) = 31.90, p < 0.001$) with a significantly larger percentage of correct responses to targets following *CṼ* primes. There is no difference between correct responses after *CV(N)* and *CV* primes ($\chi^2(1) = 0.45, p = 0.500$).

Since there is no correct or incorrect response in the *NEITHER* condition, it is not possible to provide an overall analysis of errors to see whether participants performed above chance. The results (illustrated in Figure 6; Std Errors can be found in Appendix B) were analysed by *PRIME VOWEL* and participants showed a significant preference for one target over the other in all three conditions. When presented with a *CṼ* (nasal vowel) prime, participants responded with a significantly greater number of *CVN* targets (63.8%) than *CVC* targets (36.2%; $\chi^2(1) = 51.16, p < 0.001$). In the *CV(N)* (nasalised vowel) prime condition, *CVC* targets (66.2%) were significantly more frequent than *CVC* targets (33.8%; $\chi^2(1) = 70.51, p < 0.001$) and when presented with a *CV* (oral vowel) prime, participants responded with *CVN* targets 57.4% of the time over *CVC* targets, which were chosen significantly less frequently (42.6%; $\chi^2(1) = 14.56, p < 0.001$).

In a logit generalised linear model, we again see a significant effect for *PRIME VOWEL* ($\chi^2(2) = 11.91, p < 0.003$). The planned comparisons show that participants display a greater bias towards the specific target vowels above after both *CV(N)* and *CṼ* primes than after *CV* primes (*CV(N)* versus *CV*: $\chi^2(1) = 11.10, p < 0.001$; *CṼ* versus *CV*: $\chi^2(1) = 5.89, p = 0.015$). There is no difference in the degree of preference of the target vowels after *CV(N)* and *CṼ* primes ($\chi^2(1) = 0.81, p = 0.368$).

3.3 Priming doublet results

For the doublet analysis, the analyses were split into two separate models – *NoCVN*, where no word with a nasalised vowel exists in Bengali, and *NoCVC*, where no word with a nasal vowel exists in Bengali.

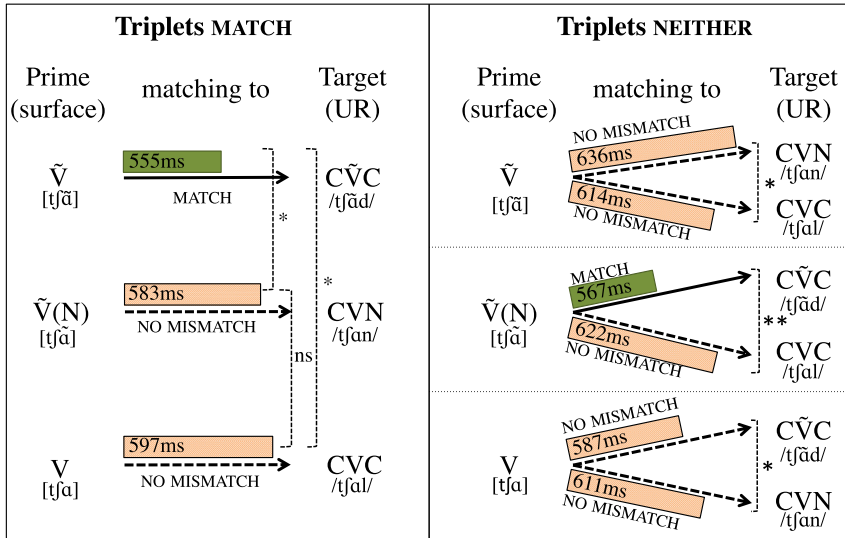


Figure 6: Predictions and reaction time results for the triplet condition

In the NoCVN condition, errors were analysed using a logit generalised linear model treating responses (error/correct) as a binomial distribution, for the fixed effect of PRIME VOWEL, with random intercepts specified for PARTICIPANTS and TARGETS. Reaction times were analysed using a linear mixed model for the fixed effect of PRIME VOWEL, with random intercepts specified for PARTICIPANTS and TARGETS.

3.3.1 Reaction time results

Reaction times of correct responses are analysed separately for the two conditions (NoCVN and NoC $\tilde{V}C$).

The linear mixed model analysis shows a different effect for PRIME VOWEL ($F(1, 1145) = 17.02, p < 0.001$) in the two sets. In the NoCVN (no contextualised nasal) set, the analysis shows significantly faster reaction times following a C \tilde{V} prime compared to a CV prime ($F(1, 750.4) = 24.36, p < 0.001$). The data in the doublet set without nasal vowels (NoC $\tilde{V}C$), however, shows no significant effect of PRIME VOWEL ($F(1, 750.4) = 0.11, p = 0.736$).

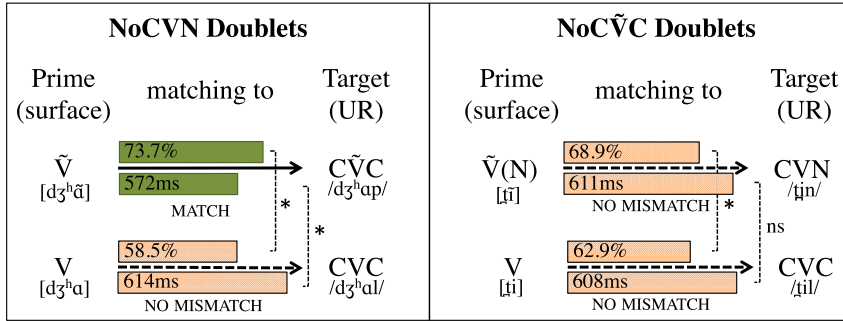


Figure 7: Reaction time results (ms) and error rates for doublet conditions.

3.3.2 Error analysis

The overall errors in the two doublet sets were 34%. The data in the doublet condition shows clearly that while listeners detect the identity match well above chance in all conditions ($\chi^2(1) = 239.10$, $p < 0.001$) and they performed equally well in both doublet sets ($\chi^2(1) = 0.01$, $p = 0.922$) there are significant differences between the different prime vowels in both reaction times and errors (cf. Figure 7; Std Errors can be found in Appendix B).

In the analysis of prime vowels in the set in which no word with a nasalised vowel exists in Bengali (NoCVN), the data shows a significant difference between the number of errors made following CVC primes versus CV primes ($\chi^2(1) = 30.82$, $p < 0.001$). Reactions to targets preceded by CV primes were substantially more accurate (26.3% errors) than those observed when primes contained an oral vowel (CV; 41.5% errors).

In the NoCVC set, where a corresponding target with a nasal vowel did not exist in Bengali, participants made significantly fewer errors ($\chi^2(1) = 4.56$, $p = 0.033$) when reacting to primes with nasalised vowels before a nasal consonant (CV(N); 31.1%) than when the targets followed oral vowel primes (CV; 37.1%). Furthermore, despite the significantly higher error rates for targets following CV primes, participants still performed significantly above chance ($\chi^2(1) = 38.16$, $p < 0.001$).

4 Discussion

The purpose of the present study was twofold: to determine whether the evidence supports an account proposing [NASAL] as a privative or an equipollent feature and to investigate whether listeners' responses were guided by the information

specified in the lexicon or whether they react to surface nasality. We have used a three-way matching process to account for the relationship between the cues from the signal and features in the lexicon. If all features are fully specified, only an exact match or a mismatch are possible. However, if a feature is not specified, for instance if we assume a lack of specification as non-nasal for oral and contextually nasalised vowels, the relationship between the surface cues and representation invokes no-mismatch rather than a genuine mismatch or, indeed, a match if they were specified as [NASAL]. Overall, the data provides strong support for a privative or monovalent representation of nasality and thus for the lack of specification of oral vowels as well as contextually nasalised vowels (by regressive assimilation from a following nasal consonant) as nonnasal. While there have been diverging proposals in the literature and different levels of specification have been put forward for different languages (see discussion in 1.1), the data collected here is best explained by the predictions made by an underspecification account such as the FUL model (Lahiri and Reetz 2002, 2010; cf. Section 1.4).

As illustrated in Figure 2, models such as this would predict that only [NASAL] is represented and extracted, and thus the only real match is either the identity match between a CVC prime and a CVC target or that found between a CVN prime and a CVC target, since the surface nasality in the signal of a CVN word can be mapped onto the specified feature [NASAL]. FUL makes clear predictions for reaction times (Lahiri and Reetz 2002) for match and no-mismatch conditions and we would thus expect listeners to respond significantly faster to match conditions than to those which result in a no-mismatch. These predictions are borne out in the reaction time data for both the triplet and doublet experiments.

In the triplet MATCH analyses (which only included correct trials), the only condition which results in significantly faster reaction times between identity primes and targets is that of CVC primes, which is also the only condition which constitutes a real match between acoustic signal and underlying representation. There is no significant difference between the RTs after CVC and CVN primes, both no-mismatch conditions, and RTs are significantly slower in both than those to CVC targets after a matching fragment prime. This data suggests that oral vowels in CVC words are not specified either as nonnasal or as oral, as otherwise there should have been no difference between response latencies after CVC and CVC primes and their respective identity targets since both primes would result in a match with their underlying lexical representation of [+nasal] and [-nasal] respectively. The error data in the MATCH condition shows an identical pattern, with participants providing correct responses significantly more frequently after CVC primes than after CVN or CVC primes, while there is no difference in the error rates between responses following the latter two primes.

In the case of the CVN items, if one assumes an equipollent set of features e.g. [\pm nasal], two competing hypotheses are conceivable concerning their representation: (a) CVN is represented as [$-$ nasal] (e.g. Chomsky and Halle 1968) or (b) CVN is represented as [$+$ nasal] (e.g. Ohala and Ohala 1995). If (a) was the case, we should find a mismatch between a CVN prime and a CVN target as the nasality in the signal would mismatch with a non-nasal specification, which would result in significantly longer reaction times than for a CVC prime and CVC target since this would be a match. This is not borne out by the results (cf. Figure 6) as there is no significant difference between the CVC and CVN conditions in the triplet *MATCH* analyses, which suggests that the relationship between acoustic input and representation is of a similar nature in both cases. In the case of hypothesis (b), which would correspond to our surface predictions (cf. Figure 1), all three prime-target combinations should result in identical reaction times since all constitute complete matches between the surface cues and the representations. However, this is clearly also not the case.

The *NEITHER* triplet condition shows similar patterns in the RT results when listeners are confronted with a nasalised vowel from a CVN prime: $C\check{V}C$ targets are chosen significantly faster than CVC targets since the former are a match, as the nasality in the signal is matched with the stored feature [*NASAL*] in the $C\check{V}C$ target while the relationship with the CVC target, which is underspecified, is a no-mismatch case (see Figure 2, middle). After $C\check{V}C$ and CVC primes, where both response options available result in a no-mismatch, the data shows no significant differences between RTs. Thus, the results do not suggest that the CVN words are represented as specified for [*NASAL*] nor is there any support for the view that the CVC words are represented as [$-$ nasal]. If the CVN words, for example, were represented as [*NASAL*], we would expect listeners to respond faster to CVN words after a fragment prime from a $C\check{V}C$ word since the surface nasality in the prime should match the stored feature [*NASAL*]. However, these RTs are significantly slower than those for CVC targets. In the condition with CVC fragment primes, there is no complete match available with either set of predictions, since the choices are CVN and $C\check{V}C$ targets. Again, it is not the case that oral vowels are being matched, because the $C\check{V}C$ targets are responded to faster. The degree of priming is significantly greater after $C\check{V}C$ primes than after either of the other primes, which further supports the facilitatory effect of the match with the specified feature in the underlying representation.

In the NoCVN doublet set, where listeners are presented with a choice between $C\check{V}C$ and CVC targets, they respond significantly faster to matching targets preceded by $C\check{V}C$ primes. When presented with [bã] from [bãḍʰ], listeners make an identity match significantly faster than when they are presented with [ba] from [baḍ]. Assuming the $C\check{V}C$ word has a specified nasal vowel, this constitutes

a better match for its identity fragment, while the vowel in the CVC is not specified and is a no-mismatch for its CV fragment although their surface relationship is identical. If we assumed a binary feature, with an oral specification for the oral vowel, we should expect to observe no difference in the reaction times since both conditions would constitute a direct match both on the surface (phonetically) and in terms of the represented features (phonologically) and should therefore elicit similar reaction times. The NoC \check{V} C set, where both vowels are unspecified, provides further support for an underspecification account (see Figure 2) since here the reaction times do not differ significantly but latencies after both CVC and CVN primes are considerably slower than those after the C \check{V} C primes in the NoCVN set.

If frequency were the crucial determining factor, which is often assumed in exemplar-based models, and it were the case that \check{V} is more frequently heard in the context of C \check{V} C than CVN, one could argue that frequency guides listeners' choices, with shorter response latencies for C \check{V} C targets after C \check{V} primes. However, if that were the case, as the oral vowel only ever occurs in the CVC context, CV primes and CVC targets should result in the fastest RTs, which is not what the data shows. This is not to say that frequency plays no role, since we also observe effects which could be attributed to frequency – when, for instance, the relationship between the prime and two targets is a no-mismatch in both cases and thus the phonological specifications do not guide responses (e.g. CV(N) prime and CVC or CVN target). However, overall the phonological representations play a critical role in listeners' perception. A further consideration could be the number of neighbours: for instance, C \check{V} C could be argued to have fewer neighbours and therefore result in faster response latencies. However, CVN words will have even fewer neighbours since they are constrained by the number of possible coda consonants (Bengali has three nasal consonants), while C \check{V} segments could theoretically be followed by up to 21 different oral consonants. Thus, while again this might contribute, it does not provide a comprehensive explanation of our data, as can be seen from the patterns of frequency data in Table 1, which do not correlate with the patterns observed in the data.

It seems that listeners are indeed using the lexical representation in the processing of nasal vowels. This can be seen particularly clearly in the reaction time data. However, the error data from the triplet NEITHER condition as well as the doublet conditions show that listeners evidently also take salient surface cues into account. In the NEITHER condition participants choose a C \check{V} C target if there is any nasality present in the signal regardless of whether this stems from a C \check{V} or CV(N) prime. If no C \check{V} C target is available, listeners choose a CVN word over a CVC word. The doublet data shows a similar tendency with identity matches being significantly more accurate after C \check{V} and CV(N) primes despite the fact that only the C \check{V} C

prime-target combination results in a match. Thus, even though a $C\tilde{V}$ prime and a CVN target result in a no-mismatch, listeners choose the CVN target as often as they chose the $C\tilde{V}C$ target in the match condition (CV(N) prime \sim $C\tilde{V}C$ target) since there is no $C\tilde{V}C$ target available in the No $C\tilde{V}C$ doublet set.

5 Conclusions

The data collected in this study supports the status of $[_{NASAL}]$ as a privative feature, whereby nasal vowels are specified as $[_{NASAL}]$ while oral vowels and contextually nasalised vowels are not specified. We established that there was no significant difference in the acoustic degree of nasality (measured using the Q-factor of the bandwidth) between underlyingly nasal and contextually nasalised vowels (cf. Table 2). The data shows that, even in a language like Bengali where nasality in vowels is used contrastively, it is not necessary to encode predictable variation in nasality, and that listeners are sensitive to lexical representations and their processing of auditory speech signals is influenced by what is provided in the lexicon. If the listener hears any nasality on the vowel, either from an underlyingly nasal or via a following nasal consonant, and if there is a lexical choice of a specified nasal vowel, then the listener overwhelmingly chooses the underlyingly nasal. If, however, nasality is perceived and there is no nasal vowel in the lexical representation, then surface nasality wins. This can only be seen in the error data and not in the reaction times. In terms of the reaction time for the triplets, we see that if a nasal vowel prime is heard and the choice is between a CVN and a CVC target, correct reactions to CVN words are slower than those to CVC words. Had the CVN word been specified as $[+nasal]$, these would have been chosen faster than the CVC words. Thus, lexical representations as reflected by reaction times suggest clearly that only privative $[_{NASAL}]$ is specified and governs the choice of responses. We provide an overview of our crucial findings below:

- Lexical representations are utilised by listeners in processing and recognition, and a match facilitates processing.
- Reaction time data across experiments reflects the underlying representation and illustrates that real matches between the signal in the prime and the representation of the target lead to faster reaction times (cf. Figure 6).
- Error analyses show that any nasality in the signal predisposes the listener to choose a target with an underlyingly nasal vowel ($C\tilde{V}$ and CV(N) to $C\tilde{V}C$), unless such a target is unavailable, in which case surface matching of nasality is used ($C\tilde{V}$ prime to CVN target).
- Both reaction time and error data provide clear support for an underspecification account of nasality with $[_{NASAL}]$ as a privative feature.

Appendix A: Full list of stimuli

1. Triplets

	CVC			CVN			CVC		
1	tʃād	চাঁদ	moon	tʃan	চান	bath	tʃal	চাল	unboiled rice
2	kāʃ	কাঁচ	glass	kan	কান	ear	kal	কাল	yesterday/tomorrow
3	pāʃ	পাঁচ	five	pan	পান	betel leaf	paʃ	পাত	place setting
4	bʰār	ভাঁড়	clay cup	bʰan	ভান	pretence	bʰaʃ	ভাত	cooked rice
5	dāʃ	ডাঁট	arrogance	dʌn	ডান	right	dʌl	ডাল	lentils
6	kʰūʃ	খুঁত	flaw	kʰun	খুন	murder	kʰur	খুর	hoof
7	tāʃ	তাঁত	loom	ʃan	তান	tune	ʃaʃ	তাস	cards
8	dʰāʃ	ধাঁচ	style, shape	dʰan	ধান	grain (rice)	dʰar	ধার	sharp
9	ʃōk	শৌঁক	smell	ʃon	শোন	listen	ʃok	শোক	grief
10	fāʃ	শাঁখ	conch shell	fʌn	শাণ	whetstone	fʌʃ	সাত	seven
11	gōʃ	গঁদ	glue	gʌm	গম	wheat	gʌʃ	গৎ	musical note
12	gʰūʃ	ঘুঁষ	bribe	gʰum	ঘুম	sleep	gʰur	ঘুর	whirl
13	dāt	দাঁত	tooth	dʌn	দান	donation	dʌʃ	দাস	servant
14	bāʃ	বাস	bamboo	ban	বান	flood	baʃ	বাদ	left out

2. Doublets (no nasal vowel)

	CVN			CVC		
1	lom	লোম	body hair	lobʰ	লোভ	greed
2	tʃen	চেঁন	chain	tʃek	চেঁক	cheque
3	gan	গান	song	gal	গাল	cheek
4	kʰam	খাম	envelope	kʰap	খাপ	step
5	dʒan	জান	astrologer	dʒaʃ	জাত	race
6	gʰam	ঘাম	drowsiness	gʰaʃ	ঘাট	bank
7	tʃun	চুঁণ	lime	tʃul	চুল	hair
8	dʌn	দিন	day	dʌk	দিক	direction
9	ʃan	টান	pull	ʃak	টাক	bald
10	tʃin	তিন	three	tʃil	তিল	sesame seed
11	ʃoŋ	সঙ	clown	ʃokʰ	সখ	desire, wish
12	roŋ	রঙ	colour	roʃ	রস	juice
13	dʰum	ধুম	smoke	dʰup	ধূপ	incense
14	gun	গুঁ	good quality	gul	গুল	charcoal ball

3. Doublets (no nasal consonant)

	CVC			CVC		
1	l̥oṭ	লৌট	lips	l̥ok	লোক	knock
2	ɕʰaṭ	ঝাট	sweeping with a broom	ɕʰal	ঝাল	spicy hot
3	ʧ̥uṭ	ছুট	needle	ʧ̥uṭ	ছুট	run, escape
4	ʧ̥aṭ	ছাট	mould, cast	ʧ̥aṭ	ছাত	terrace
5	ʧ̥oṭ	ছৌড়	throw	ʧ̥oṭ	ছোট	run
6	ɕʰāp	ঝাঁপ	leap	ɕʰaṭ	ঝাড়	bush, shrub
7	ɕʰoṭk	ঝোঁক	wish, inclination	ɕʰol	ঝোল	broth
8	pʰaṭ	ফাঁস	knot, noose	pʰaṭ	ফাগ	red powder to play with
9	ʃākʰ	শাঁখ	conch shell	ʃak	শাক	greens, spinach
10	hāṭ	হাঁস	duck	haṭ	হাত	hand
11	ʧ̥oṭ	চৌচ	fibre, splinter	ʧ̥or	চোর	thief
12	gōṭʰ	গৌতি	knot	git	গীত	song
13	pūṭ	পুঁজ	pus, infection	pur	পুর	city
14	ʃūr	সূর	trunk of an animal	ʃud	সুদ	interest (payment)

Appendix B: Standard errors for reaction time and error analyses

Table 1: Standard errors for error analyses of the triplet set.

MATCH						
Prime	CVC		CVN	CVC		
Probability of correct	68.8%		57%	58.3%		
(StdError)	(1.26)		(1.35)	(1.35)		
NEITHER						
Prime	CVC		CVN	CVC		
chosen	CVC	CVN	CVC	CVC	CVC	CVN
Probability of chosen	36.2%	63.8%	66.2%	33.8%	42.6%	57.4%
(StdError)	(1.86)	(1.86)	(1.82)	(1.82)	(1.91)	(1.91)

Table 2: Triplet reaction time results (combined RT for correct responses and identity responses by context (i.e. the other available prime) in the MATCH condition and combined responses as well as RT per chosen target in the NEITHER condition).

MATCH						
Prime	CVC		CVN		CVC	
context of response	CVN	CVC	CVC	CVC	CVC	CVN
RT (ms)	555	556	569	597	608	586
(StdError)	(13.14)	(12.93)	(13.73)	(13.05)	(13.40)	(13.10)
RT (ms) overall	555		583		597	
(StdError)	(11.26)		(11.47)		(11.38)	
NEITHER						
Prime	CVC		CVN		CVC	
chosen	CVC	CVN	CVC	CVC	CVC	CVN
RT (ms)	614	636	567	622	587	611
(StdError)	(14.27)	(13.35)	(13.32)	(14.45)	(13.95)	(13.48)
RT (ms) overall	625		594		599	
(StdError)	(12.96)		(12.98)		(12.91)	

Table 3: Standard errors for reaction time and error analyses of the doublet sets.

NoCVN DOUBLETS		
Prime	CVC	CVC
RT (ms) overall	572.14	614.39
(StdError)	(11.29)	(11.64)
Probability of correct	73.7%	58.5%
(StdError)	(3.70)	(3.88)
NoCVC DOUBLETS		
Prime	CVN	CVC
RT (ms) overall	611.23	608.40
(StdError)	(12.29)	(12.42)
Probability of correct	68.9%	62.9%
(StdError)	(3.47)	(3.74)

References

- Beddor, Patrice Speeter. 1993. The perception of nasal vowels. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, Nasalization, and the Velum*. Phonetics & Phonology, vol. 5, 171–196. London: Academic Press.
- Beddor, Patrice Speeter. 2007. Nasals and nasalization: The relation between segmental and coarticulatory timing. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany.
- Beddor, Patrice Speeter & Winifred Strange. 1982. Cross-language study of perception of the oral-nasal distinction. *The Journal of the Acoustical Society of America* 71. 1551–1561.
- Beddor, Patrice Speeter, Kevin B. McGowan, Julie E. Boland, Andries W. Coetzee & Anthony Brasher. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133. 2350–2366.
- Berger, Michael A. 2007. *Measurement of vowel nasalization by multi-dimensional acoustic analysis*. Rochester, New York, MA: University of Rochester.
- Bjuggren, Gunnar & Carl Gunnar Michael Fant. 1964. The nasal cavity structure. *Speech Transmission Laboratory – Quarterly Progress and Status Report* 5(4). 5–7.
- Boersma, Paul & David Weenink. 2011. *PRAAT: Doing Phonetics by Computer (ver. 5.2.26)*. Amsterdam, The Netherlands: Institute for Phonetic Sciences.
- Butcher, Andrew. 1976. The influence of the native language on the perception of vowel quality. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* 6. 1–137.
- Byrne, William, Michael Finke, Sanjeev Khudanpur, John McDonough, Harriet Nock, Michael Riley, Murat Saraçlar, Charles Wooters & George Zavalagkos. 1998. Pronunciation modelling using a hand-labelled corpus for conversational speech recognition. In *Proceedings of ICASSP*, vol. 98, 313–316.
- Carignan, Christopher. 2018. Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *The Journal of the Acoustical Society of America* 143. 2588–2601.
- Chatterji, Suniti Kumar. 1926. *Origin and Development of the Bengali Language*. Calcutta, India: Calcutta University Press (reprinted by Rupa & Co., Calcutta, 1975).
- Chen, Marilyn Y. 1996. *Acoustic Correlates of Nasality in Speech*. Cambridge, MA: Massachusetts Institute of Technology PhD.
- Chen, Marilyn Y. 1997. Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America* 102. 2360–2370.
- Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York, NY: Harper & Row.
- Clements, G. Nick. 2001. Representational economy in constraint-based phonology. In Tracy Alan Hall (ed.), *Distinctive feature theory*, 71–146. Berlin: Mouton de Gruyter.
- Cohn, Abigail C. 1993. Nasalisation in English: Phonology or phonetics? *Phonology* 1. 43–81.
- Cooper, Nicole, Anne Cutler & Roger Wales. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45(3). 207–228.
- Dang, Jianwu, Kiyoshi Honda & Hisayoshi Suzuki. 1994. Morphological and acoustical analysis of the nasal and the paranasal cavities. *The Journal of the Acoustical Society of America* 96. 2088–2100.

- Dang, Jianwu & Kiyoshi Honda. 1996. Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation. *The Journal of the Acoustical Society of America* 100. 3374–3383.
- Dash, Niladri Sekhar & Bidyut Baran Chaudhuri. 2001. A corpus-based study of the Bengali language. *Indian Journal of Linguistics* 20(1). 19–40.
- Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.
- Elman, Jeffrey L. & James L. McClelland. 1986. Exploiting lawful variability in the speech wave. In Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and variability of speech processes*, 360–380. Hillsdale, NJ: Erlbaum.
- Fant, Carl Gunnar Michael. 1960. *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton & Co.
- Ferguson, Charles A. & Munier Chowdhury. 1960. The phonemes of Bengali. *Language* 36(1). 22–59.
- Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics* 14. 3–28.
- Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1). 110–125.
- Gaskell, M. Gareth & William D. Marslen-Wilson. 1997. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12. 613–656.
- Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP*, 517–520.
- Grosjean, François. 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28. 267–283.
- Hajek, John. 2013. Vowel Nasalization. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/10> (accessed 28 February 2022).
- Halle, Morris, Bert Vaux & Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31. 387–444.
- Hattori, Shiro, Kengo Yamamoto & Osamu Fujimura. 1958. Nasalization of vowels in relation to nasals. *The Journal of the Acoustical Society of America* 30. 267–274.
- Hawkins, Sarah & Kenneth N. Stevens. 1985. Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels. *The Journal of the Acoustical Society of America* 77. 1560–1575.
- House, Arthur S. & Kenneth N. Stevens. 1956. Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders* 21. 218–232.
- Huffman, Marie K. 1990. *Implementation of Nasal: Timing and Articulatory Landmarks*, vol. 75. UCLA PhD thesis.
- Hyman, Larry. 1973. The feature [Grave] in phonological theory. *Journal of Phonetics* 1. 329–337.
- Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullennix (eds.), *Talker variability in speech processing*, 145–165. San Diego, CA: Academic.
- Jongman, Allard, Joan A. Sereno, Marianne Raaijmakers & Aditi Lahiri. 1992. The phonological representation of [voice] in speech perception. *Language and Speech* 35(1, 2). 137–152.
- Klaiman, Miriam Holly & Aditi Lahiri. 2018. Bengali. In Bernard Comrie (ed.), *The World's Major Languages*, 427–446. London: Routledge.
- Krämer, Martin. 2017. Is vowel nasalisation phonological in English? A systematic review. *English Language and Linguistics* 23(2). 405–437.

- Lahiri, Aditi. 2018. Predicting universal phonological features. In Larry Hyman & Frans Plank (eds.), *Phonological Typology*, 229–272. Berlin: Mouton de Gruyter.
- Lahiri, Aditi & William D. Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38. 245–294.
- Lahiri, Aditi & William D. Marslen-Wilson. 1992. Lexical processing and phonological representation. In Robert D. Ladd & Gerard Docherty (eds.), *Papers in Laboratory Phonology II: gesture, segment, prosody*, 229–254. Cambridge: Cambridge University Press.
- Lahiri, Aditi & Henning Reetz. 2002. Underspecified recognition. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology VII*, 637–677. Berlin, Germany: Mouton de Gruyter.
- Lahiri, Aditi & Henning Reetz. 2010. Distinctive Features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38. 44–59.
- Lieberman, Alvin M. & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21. 1–36.
- Lindqvist-Gauffin, Jan & Johan Sundberg. 1976. Acoustic properties of the nasal tract. *Phonetica* 33. 161–168.
- Maeda, Shinji. 1982. The role of the sinus cavities in the production of nasal vowels. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 911–914. <https://10.1109/ICASSP.1982>.
- Maeda, Shinji. 1983. *Acoustic cues of vowel nasalization: A simulation study*, VII, 25–36. Paris: Recherches/Acoustique Centre National d'Etudes des Telecommunications (CNET).
- Malécot, André. 1960. Vowel nasality as a distinctive feature in American English. *Language* 36. 222–229.
- Marslen-Wilson, William D. 1987. Functional parallelism in spoken word recognition. *Cognition* 25. 71–102.
- Marslen-Wilson, William D. 1989. Access and integration: Projecting sounds onto meaning. In William D. Marslen-Wilson (ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, 148–172. Cambridge, MA: MIT Press.
- Marslen-Wilson, William D. 1993. Issues of process and representation in lexical access. In Gerry T. M. Altmann & Richard Shillcock (eds.), *Cognitive models of speech processing: The Second Sperlonga Meeting*, 187–210. Hillsdale, NJ: Erlbaum.
- Marslen-Wilson, William D. & Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10. 29–63.
- McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 43. 84–108.
- McClelland, James L. & Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 10. 1–86.
- McQueen, James M. 2005. Speech perception. In Koen Lamberts & Robert Goldstone (eds.), *The Handbook of Cognition*, 255–275. London: Sage.
- Norris, Dennis. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52. 189–234.
- Norris, Dennis & James M. McQueen. 2008. Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review* 115(2). 357–395.
- Norris, Dennis, James M. McQueen & Anne Cutler. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23. 299–325.
- Ohala, John J. & Manjari Ohala. 1995. Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology*

- and Phonetic Evidence. *Papers in Laboratory Phonology IV*, 41–60. Cambridge: Cambridge University Press.
- Pallier, Christophe, Anne Christoph & Jacques Mehler. 1997. Language-specific listening. *Trends in Cognitive Sciences* 1. 129–132.
- Pallier, Christophe, Angels Colomé & Nuria Sebastian-Gallés. 2001. The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science* 12. 445–449.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology*, vol. 7, 101–139. Berlin: De Gruyter.
- Pruthi, Tarun. 2007. *Analysis, Vocal-Tract Modeling and Automatic Detection of Vowel Nasalization*. Maryland: University of Maryland PhD.
- Pruthi, Tarun, Carol Y. Espy-Wilson & Brad H. Story. 2007. Simulation and analysis of nasalized vowels based on magnetic resonance imaging data. *The Journal of the Acoustical Society of America* 121. 3858–3873.
- Reetz, Henning & Achim Kleinmann. 2003. Multi-subject hardware for experiment control and precise reaction time measurement. In Maria-Josep Solé, Daniel Recasens & Joachim Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1489–1492. Barcelona.
- Scharinger, Matthias & Aditi Lahiri. 2010. Height differences in English dialects: Consequences for processing and representation. *Language and Speech* 53(2). 245–272.
- Solé, Maria-Josep. 1992. Phonetic and phonological processes: The case of nasalization. *Language and Speech* 35. 29–43.
- Solé, Maria-Josep. 1995. Spatio-temporal patterns of velo-pharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38. 1–23.
- Stevens, Kenneth N. 1998. *Acoustic Phonetics*. Cambridge, MA: The MIT Press.
- Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America* 111. 1872–1891.
- Styler, Will. 2017. On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America* 142. 2469–2482.
- ‘t Hart, Johan & Antonie Cohen. 1964. Gating techniques as an aid in speech analysis. *Language and Speech* 7. 22–39.
- Trigo, R. Lorenza. 1993. The inherent structure of nasal segments. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, Nasalisation, and the Velum*. Phonetics & Phonology, vol. 5, 369–400. New York: Academic Press.
- Vampola, Tomáš, Jaromír Horáček, Vojtěch Radolf, Jan G. Švec & Anne-Marie Laukkanen. 2020. Influence of nasal cavities on voice quality: Computer simulations and experiments. *The Journal of the Acoustical Society of America* 148(5). 3218–3231.
- Wright, James T. 1986. The behaviour of nasalized vowels in perceptual vowel space. In John J. Ohala & Jeri J. Jaeger (eds.), *Experimental phonology*, 45–67. New York: Academic Press.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/phon-2022-2017>).