PERCEPTION AND REPRESENTATION OF BENGALI NASAL VOWELS

Sandra Kotzor[1], Adam Roberts[2], Allison Wetterlin[1, 3] & Aditi Lahiri[1]

[1]University of Oxford, [2]Nanyang Technological University, [3]Universitet i Agder
sandra.kotzor@ling-phil.ox.ac.uk

ABSTRACT

The paper addresses the question of native speakers' online awareness and perceptual use of phonetic nasalisation by examining surface nasalisation in two types of surface vowels in Bengali: underlying nasal vowels and nasalised vowels before a nasal consonant. In a cross-modal forced-choice experiment, we investigate the hypothesis that only unpredictable nasalisation is represented and that this sparse representation governs listeners' interpretation of vowel nasality. Auditory primes consisting of CV segments of monosyllabic CVC words containing either nasal vowels ([cã] for *cãd*), oral vowels ([ca] for *cal*) or nasalised oral vowels ([ca(n)] for *can*) preceded visual full-word targets.

Faster reaction times and fewer errors are observed after nasal vowel primes compared to both oral and nasalised vowel primes. This indicates that nasal vowels are specified for nasality and lead to faster recognition compared to the oral vowel conditions, which are underspecified and thus cannot be perfectly matched with incoming signals.

Keywords: perception, lexical representation, nasal vowels, phonology, underspecification

1. INTRODUCTION

The oral-nasal contrast for vowels is well established and nasal vowels are considered marked. For instance, languages which have nasal vowels also invariably have the corresponding oral vowels but not vice versa. Within a monovalent feature system, nasal vowels are generally assumed to be represented with the feature [NASAL] while oral vowels are unspecified [1, 7]. In contrast, binary features would mark both oral and nasal vowels, namely as [-NASAL] and [+NASAL] respectively. Nasality in vowels can stem from two different sources: either the vowel is underlyingly nasal, or it is underlyingly oral and becomes nasal by regressive assimilation from a neighbouring nasal segment.

In Bengali, we find underlying nasal vowels ([cãd] 'moon') corresponding to each of the seven underlying oral vowels in the language ([cal] 'unboiled rice' [4]), which can also surface as nasalised vowels in the context of a following nasal consonant ([cãn] 'bath').

An underspecification account of nasality [7] would propose that the nasal vowel (NV) is specified for [NASAL] while both the oral vowel (OV) and the oral vowel preceding a nasal consonant (NC) are not specified despite the fact that the surface realisation of the vowel in the latter case is nasalised. This results in an ambiguous surface realisation of NV and NC vowels which have different underlying representations.

In an off-line gating task investigating nasal vowels in Bengali, Lahiri & Marslen-Wilson [5] found an asymmetry in the responses to nasal and oral vowels. Subjects were asked to provide the complete word (unconstrained) after hearing increasing increments of monosyllabic Bengali words. At the gate at the end of the vowel, OV input never resulted in NV responses, which may be due to a mismatch between the acoustic signal and the lexical representation since NV items are specified for nasality. OV and NC items, however, are underspecified, and any incoming signal creates a no-mismatch condition. In these cases participants were guided by the phonetic characteristics of the vowel, and mainly provided OV and NC responses (in proportion to CVN words in the lexicon) to OV primes and both NV and NC responses to NC primes.

Ohala & Ohala [8] replicated Lahiri & Marslen-Wilson's gating study with minor modifications but proposed that 'at best, [Lahiri & Marslen-Wilson's data] are compatible with both the UR [underspecified representation] and SR [surface representation] hypotheses', since most of their findings could be explained by a surface representation approach.

The present research uses a cross-modal (audio-visual) form priming task with a forced-choice response paradigm (cf. [3]) to investigate the processing and thus the representation of nasality in the context of vowels in an online experiment.

If, as we predict, nasal vowels (NV) are specified as [NASAL] while oral vowels (OV) and vowels nasalised due to regressive assimilation

(NC) are unspecified for nasality, we should observe shorter response latencies as well as lower error rates after NV primes than after OV and NC primes. This is due to differences in representation: NV signals result in a match with [NASAL] in the recognition process while OV and NC result in a no-mismatch, which should yield responses with longer latencies than the match condition [6].

2. MATERIALS AND DESIGN

2.1. Stimuli

2.1.1. Primes

All primes consisted of the CV fragment of common monosyllabic CVC words of Bengali and the stimuli were split into three sets of minimal pairs: one set of 14 triplets and two sets of 14 doublets (cf. Table 1). In the triplet set, all possible variations of nasality were attested in the language: oral vowels ([cal] 'unboiled rice'), nasal vowels ([cãd] 'moon') and nasalised oral vowels due to a following nasal ([cãn] 'bath'). In the first doublet set (NoNV), no words with the same consonant sequence containing a nasal vowel exist so the two primes in this set were oral vowels ([til] 'black seeds') and nasalised oral vowels due to a following nasal ([tin] 'three'). The second doublet set (NoNC) contained only oral vowels ([j^hal] 'spicy') and underlying nasal vowels ([j^hãp] 'jump') but no minimally different CV pairs with nasalised oral vowels exist. All primes were CV fragments of the words presented and were truncated to include the complete duration of the vowel but no consonantal information.

Table 1: Prime-target combinations

	Doublets (NoNV)		Doublets (NoNC)	
Prime	**[ti(n)]**	**[ti]**	**[j^hã]**	**[j^ha]**
Target (1)	tin - til	til - tin	j^hal - j^hãp	j^hãp - j^hal
Target (2)	til - tin	tin - til	j^hãp - j^hal	j^hal - j^hãp

	Triplets		
Prime	**[cã]**	**[ca(n)]**	**[ca]**
Target (Block 1)	can - cãd	can - cãd	can - cãd
	cal - can	cal - can	cal - can
	cãd - cal	cãd - cal	cãd - cal
Target (Block 2)	cãd - can	cãd - can	cãd - can
	can - cal	can - cal	can - cal
	cal - cãd	cal - cãd	cal - cãd

2.1.2. Targets

Targets were the full words corresponding to the fragment primes. In the case of the triplets, two of the possible three primes were presented and in one third of the cases neither of the two targets matched the prime (cf. Table 1).

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, digitised and the volume equalised using PRAAT [2].

2.2. Procedure

The experiment was conducted at Gokhale Memorial Girls' College in Kolkata, India. Participants were tested in groups of a maximum of 16 in a quiet and darkened room. The auditory primes were played through individual closed-ear headphones (SONY MDR110 LP) and visual targets were presented from a MacBook Pro in Bengali script (font size 36) and were projected onto a screen. Subjects responded via custom-made individual two-button boxes with the left button corresponding to the left target and the right to the right target. Subjects were right-handed and thus used their dominant hand to indicate a match of the prime and the target displayed on the right.

Trials were separated into two blocks. Each block consisted of 182 pseudo-randomised trials presented in a different order in each block. 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 each combination of targets occurred once in each block. Side of presentation (left vs. right) of the target words was randomised and counterbalanced across blocks. Targets were displayed for 800ms with an ISI of 0ms and the interval between trials was 1500ms. Subjects heard a beep before each prime followed by a 200ms silence and a sequence of three beeps after every fourteen trials.

3. RESULTS

3.1. Analysis

Overall, two participants' data was discounted due to a malfunctioning button box and one additional subject was omitted from the analysis due to consistently fast reaction times suggesting reactions to the prime rather than the target. In addition, all reaction times ≤ 0 and reaction times outside ±2 standard deviations were excluded from the analyses.

3.2. Doublet error results

The overall errors in the two doublet sets were 38%. Thus subjects performed well above chance when identifying the matching target ($\chi^2(1)$ = 176.07, p < .001) and they performed equally well in both doublet sets ($\chi^2(1)$ = 0.10, p = .748).

In the analysis of prime vowels in the set in which no word with a nasalised vowel (NC < CVN) exists in Bengali the data shows a significant difference between the number of errors made following NV primes vs. OV primes ($\chi^2(1)$ = 19.47, p ≤ .001). The reactions to targets preceded by NV primes were much more accurate (32.3% errors) than when primes contained an oral vowel (OV; 43.2% errors).

In the NoNV set, where a corresponding target with a nasal vowel did not exist in Bengali, participants made significantly fewer errors ($\chi^2(1)$ = 4.52, p = .033) reacting to primes with nasalised vowels before a nasal consonant (35.7%) than when the targets followed oral vowel primes (41%).

The analyses also showed that despite the significantly higher error rates for targets following OV primes, participants still performed significantly above chance ($\chi^2(1)$ = 38.80, p < .001).

Table 2: Doublet results

	NoNC Set				NoNV Set			
Prime	NV		OV		NC		OV	
correct	Y	N	Y	N	Y	N	Y	N
% correct	67.7	32.3	56.8	43.2	64.3	35.7	59	41
RT (ms)	582		617		611		619	

3.3. Doublet reaction time results

The reaction times are analysed separately for the two conditions (NoNC and NoNV) using a linear mixed model analysis with subjects as a random factor and prime vowel (*Pvowel*) and correct response (*Correct*) as conditions.

The linear mixed model analysis of the correct responses in the NoNC condition shows a significant difference in the response latencies following an NV prime compared to an OV prime (F(1, 920.9) = 16.79, p < .001) with significantly faster responses following NV primes.

In the doublet set without nasal vowels (NoNV), the linear mixed model analysis shows no significant difference between response latencies following a NV prime as opposed to an OV prime (F(1, 897) = 0.95, p = .329).

3.4. Triplet error results

Due to the design of the experiment, the triplet data will be analysed in two separate conditions: MATCH ONLY and NEITHER ONLY. MATCH ONLY contains all instances where there was an exact match to the prime available as a target (e.g. prime [cã]; targets: *can* - *cãd*). NEITHER ONLY contains the data from those trials where no identity match was available (e.g. prime [cã]; targets: *can* – *cal*).

Overall, in the *MATCH ONLY* condition, participants responded with the identity target 59.9% of the time and thus performed significantly better than chance in this task ($\chi^2(1)$ = 176.99, p < .001). In analyses by prime vowel, participants' responses were significantly more accurate than chance in all three conditions: NV ($\chi^2(1)$ = 175.49, p < .001), NC ($\chi^2(1)$ = 18.12, p < .001) and OV ($\chi^2(1)$ = 31.42, p < .001).

In a logit generalised linear model, we see a significant effect for prime vowel ($\chi^2(2)$ = 48.05, p < .001). In a planned comparison, comparing the percentages of correct responses to targets we find that there is a significant difference between responses following NV vs. NC primes ($\chi^2(2)$ = 41.61, p < .001) as well as those after NV vs. OV primes ($\chi^2(2)$ = 29.85, p < .001) with a significantly larger percentage of correct responses to targets following NV primes. There is no difference between correct responses after NC and OV primes ($\chi^2(2)$ = 0.95, p < .330).

Since there is no correct or incorrect response in the *NEITHER ONLY* condition, it is not possible to provide an overall analysis of errors to see whether participants performed above chance. The results were analysed by prime vowel and participants showed a significant preference for one target over the other in all three conditions. When they heard an NV (nasal vowel) prime, they responded with a significantly greater number of NC targets (62.6%) than OV targets (37.4%; $\chi^2(1)$ = 47.76, p < .001). In the NC (nasalised vowel) prime condition, NV targets (64.7%) were significantly more frequent than OV targets (35.3%; $\chi^2(1)$ = 65.83, p < .001) and when presented with an OV (oral vowel) prime, participants responded with NC targets 57.4% of the time over NV targets, which were chosen significantly less frequently (42.6%; $\chi^2(1)$ = 16.74, p < .001).

In a logit generalised linear model, we again see a significant effect for prime vowel ($\chi^2(2)$ = 8.75, p < .013). The planned comparisons show that participants display a greater bias towards the specific target vowels above after both NC and NV primes than after OV primes (NC vs. OV: $\chi^2(2)$ = 8.28, p = .004; NV vs. OV: $\chi^2(2)$ = 4.10, p = .043).

There is no difference in the degree of preference of the target vowels after NC and NV primes ($\chi^2(2) = 0.72$, p = .396).

Table 3: Triplet results

	MATCH ONLY					
Prime	NV		NC		OV	
correct	Y	N	Y	N	Y	N
% correct	66.9	33.1	55.5	44.5	57.3	42.7
RT (ms)	549		581		595	
	NEITHER ONLY					
Prime	NV		NC		OV	
chosen	OV	NC	NV	OV	NV	NC
% chosen	37.4	62.6	64.7	35.3	42.6	57.4
RT (ms)	619		578		594	

3.5. Triplet reaction time results

In the *MATCH* condition, a linear mixed model with subjects as a random factor shows an overall significant effect for *Pvowel* (F(2, 2640) = 39.55, p < .001). In a planned comparison between the three different prime vowels, we see significant differences between the response latencies to targets after NV vs. NC primes (t(2) = -5.95, p < .001) and NV vs. OV primes (t(2) = -8.61, p < .001) with responses after NV primes being faster than those after NC and OV primes. The difference between latencies after NC and OV primes is also significant (t(2) = -2.50, p = .013) and responses after NC primes are faster than those after OV primes.

In the *NEITHER ONLY* condition, a linear mixed model (with subjects as a random factor) was performed with the variable MATCH NEITHER (*Pvowel*/*Tvowel*) and the overall effect is significant (F(5, 2198) = 19.09, p < .001). A planned comparison between the two different options for each prime vowel resulted in the following pattern: If the prime vowel was NC, subjects reacted significantly faster when they chose an NV target compared to an OV target (t(3) = -6.36, p < .001). When the prime was an OV, subjects chose an NV target faster than an NC target (t(3) = 2.53, p = .012) and when the prime contained a NV, there was no significant difference between the OV and NC targets (t(3) = 1.68, p = .093).

4. DISCUSSION

The data in the doublet condition shows clearly that, while listeners perform above chance in all conditions, there are significant differences between the different prime vowels in both reaction times and errors. In the NoNC set, with a choice between NV and OV primes, listeners not only responded much more accurately (68% vs. 57%) to targets preceded by NV primes but their response latencies were also significantly faster. In the NoNV set, where both vowels are underspecified, we still see significantly higher accuracy for the targets after NC primes (64% vs. 59%). While this difference is smaller, the nasality present in the acoustic signal potentially serves as an added cue which aids correct matching. Interestingly, the reaction times in this set are not significantly different and both are significantly slower than the reaction times after the NV primes in the NoNC set.

The doublet data provides support for an underspecified representation through both correct responses and reaction times. However, the differences in reaction times are particularly striking and are consistent with theoretical predictions of the FUL model, where a match results in shorter latencies while a no-mismatch leads to longer reaction times [5].

The triplet data in the MATCH condition shows a similar pattern: there is no difference between the correct responses after NC and OV primes (55% and 57%) while NV primes result in a significantly larger proportion of correct responses (67%). Once again, however, it is the reaction time data which shows the pattern most clearly. Latencies after NV primes are significantly shorter than after both OV and NC primes. In addition, we also see a difference between NC and OV primes with shorter latencies for NC primes. This shows clearly that oral vowels are not specified for [-NASAL], as otherwise the reaction times for NV and OV should be the same since both signals would result in a match with the lexical representation.

The NEITHER ONLY triplet condition shows the most balanced pattern of responses after OV primes. While participants here still chose NC targets significantly more frequently than NV targets, they show significantly faster RTs to NV targets. When confronted with an NC prime, listeners chose mostly NV targets which they once again reacted to faster than to OV targets. In the NV prime condition, there was no difference in reaction time between NC and OV targets but the nasality in the prime guided listeners to a significantly larger proportion of NC responses.

Overall, this data provides strong support for the underspecification of underlyingly oral vowels as being not specified for nasality, not only through the consistently faster reaction times for match conditions but also the similarities, in response latencies in particular, between underlyingly unspecified OV and NC primes.

5. REFERENCES

[1] Archangeli, D. 1984. *Underspecification in Yawelmani phonology*. Doctoral dissertation. Cambridge, MA: MIT.

[2] Boersma, P., Weenink, D. 2011. *PRAAT: Doing Phonetics by Computer (ver. 5.2.26)*. Amsterdam, The Netherlands: Institute for Phonetic Sciences.

[3] Cooper, N., Cutler, A., Wales, R. 2002. Contraints of lexical stress on lexical access in English: evidence from native and non-native listeners. *Lang Speech*, 45(3), 207–228.

[4] Ferguson, C. A., Chowdhury, M. 1960. The phonemes of Bengali. *Language*, 36, 22–59.

[5] Lahiri, A., Marslen-Wilson, W. D. 1991. The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition, 38*, 245–294.

[6] Lahiri, A., Reetz, H. 2002. Underspecified recognition. In Gussenhoven, C., Warner, N. (eds), *Laboratory Phonology VII*. Berlin, Germany: Mouton de Gruyter, 637–677.

[7] Lahiri, A., Reetz, H. 2010. Distinctive Features: Phonological underspecification in representation and processing. *J Phonetics, 38*, 44–59.

[8] Ohala, J. J., Ohala, M. 1995. Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. In: Connell, B., Arvaniti, A. (eds), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 41–60.