DISTINGUISHING BREATHY CONSONANTS AND VOWELS IN GUJARATI

Christina M. Esposito[1], Sameer ud Dowla Khan[2], Kelly Harper Berkson[3], Max Nelson[4]

[1]Macalester College, [2]Reed College, [3]Indiana University Bloomington, [4]University of Massachusetts Amherst
esposito@macalester.edu, skhan@reed.edu, kberkson@indiana.edu, manelson@umass.edu

ABSTRACT

Acoustically and articulatorily, consonantal (Cʱ) and vocalic (V̤) breathiness can be described using many of the same correlates. In Gujarati, which has both (e.g. /baɾ/ 'twelve', /bạɾ/ 'outside', /bʱaɾ/ 'burden'), breathiness is realized through greater spectral tilt, lower periodicity in the signal, and a higher open quotient; only the timing and magnitude of these features distinguish consonantal from vocalic breathiness. We explore whether native listeners distinguish breathiness from modal voice, and whether they can further distinguish breathy vowels (e.g. /bạɾ/) from vowels following breathy-voiced consonants (e.g. /bʱaɾ/). Results from free sorting, AX discrimination, and picture matching indicate that listeners can distinguish breathy from modal voice but cannot distinguish breathy vowels from vowels following breathy consonants. This suggests that speakers produce three-way distinctions of breathiness, but perceive breathiness as a binary.

Keywords: breathiness, voice quality, perception, free sort, Gujarati.

1. INTRODUCTION

Crosslinguistically, breathiness can be described using the same acoustic and articulatory correlates regardless of association to a vowel (V̤) or to a consonant (Cʱ) [1]. And, while many languages contrast breathy phonation either on obstruents as in Hindi [2, 3], Bengali [4], and Marathi [5] or on vowels as in many Zapotec languages [6, 7, 8], few languages preserve this contrast across obstruents *and* vowels. This distinction appears to be limited to some Khoisan languages (e.g. !Xóõ [9], Ju|'hoansi [10]), White Hmong [11], and Gujarati [11].

This paper explores whether Gujarati listeners distinguish breathy Vs (CV̤ e.g. /bạɾ/ 'outside') from vowels following breathy (i.e. voiced-aspirated) Cs (CʱV e.g. /bʱaɾ/ 'burden'). Previous research [12] demonstrates that the timing and degree of acoustic cues are important in distinguishing this contrast in Gujarati: vowels following breathy consonants (CʱV) are characterized by a short initial period of intense breathiness, while breathy vowels (CV̤) have a stable, more moderate breathiness throughout. Perceptually, Gujarati speakers reliably distinguish breathy vowels from modal vowels in Gujarati stimuli [11, 12], but here we ask: can listeners also leverage the differences in timing and degree of breathiness in order to reliably distinguish breathy consonants [CʱV] from breathy vowels [CV̤]?

2. METHODS

Three tasks (free sort, AX discrimination, picture-matching identification, described in 3.1–3.3) investigated the perception of CV, CʱV, and CV̤ sequences by native Gujarati listeners. Task order was not randomized, as the identification task imposed predetermined categories on listeners. To minimize any segment-specific or gender effects, stimuli consisted of a minimal triplet (Table 1) produced by four native speakers, all women from Mumbai between the ages of 22–30.

Table 1: Stimulus list.

Breathy V	બહાર	bạɾ	'outside'
Breathy C	ભાર	bʱaɾ	'burden'
All modal	બાર	baɾ	'twelve'

Stimuli were extracted from [12], in which speakers were asked to produce (as many times as possible within a 10s window) a sentence of their own creation beginning with the stimulus word. (This method helped mitigate the effects of careful speech and spelling pronunciation on breathy Vs described in [11]; we double-checked to confirm that all breathy V tokens used in the current study were produced as monosyllabic [bạɾ] and never as disyllabic [bəhaɾ].) Two repetitions of each stimulus were used, for a total of 24 tokens (3 stimuli X 2 reps X 4 talkers). Six native Gujarati listeners participated, four male listeners in their mid-20s and two female listeners (one 25 yrs, one 52 yrs).

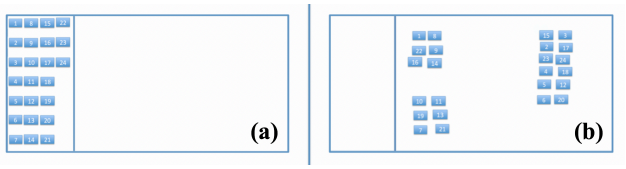
3. RESULTS

3.1. Free sort task

The free sort task [13] investigated whether listeners independently proposed three target categories ([baɾ],

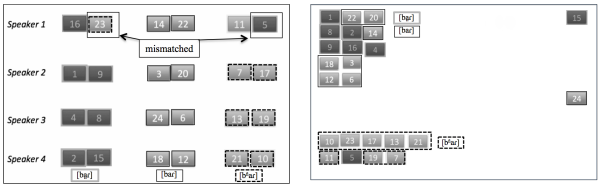
[bʱaɾ], [bạɾ]) when presented with a screen containing 24 numbered icons arranged randomly (Fig. 1a) and asked to categorize them by arranging them into groups (see sample outcome in Fig. 1b). Icons corresponded to one of the 24 audio stimuli, and played when clicked.

Figure 1: Free sort setup (a), sample outcome (b)



To avoid experimenter-imposed biases, listeners had absolute freedom over how to categorize items and how many categories to propose, and so a purely descriptive report of the outcomes is most informative. The three response patterns included (i) pairing token and speaker, (ii) separating breathy Cs from all other tokens, and (iii) groupings that were less interpretable.

Figure 2: Re-coded outcomes for two response patterns. Two listeners grouped by speaker and token-type with high accuracy (a); two created one group of breathy consonant items and a second group of plain and breathy vowel tokens (b).



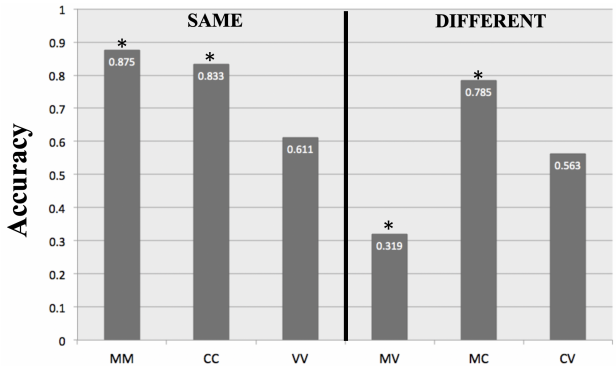
(a) (b)

Listeners 1 and 2 paired stimuli by both token and speaker, yielding 12 groups (Fig. 2a); Listener 1 was highly accurate in this pattern, Listener 2 less so. Listeners 3 and 4 formed two unique groups (see Fig. 2b), one a well-defined [bʱaɾ] category and the other combining [baɾ] and [bạɾ]. This result suggests overlap in the fully modal [baɾ] and breathy vowel [bạɾ] categories. Listener 5 created three groups, possibly intended to represent the three categories of stimuli: each consisted of a majority of one type of stimuli, but all were mixed and contained at least one member of each of the three stimuli types. The most consistent group was breathy consonants [bʱaɾ], indicating that these are the least confusable type of stimuli. Listener 6 created seemingly unstructured groups, highlighting that problems can arise in a task with so few guidelines.

3.2. Discrimination task

The discrimination task probed how accurately listeners can distinguish pairs of target words. In one sense, this task most directly addresses the issue of perceiving differences between CV, CʱV, and CV̰ sequences: while in other perceptual tasks listeners may categorize stimuli first and then compare categories rather than stimuli, a discrimination task encourages listeners to compare the stimuli directly [15]. Items were presented in a classic AX task. In the trials, participants heard two of the 24 stimuli in succession and indicated whether the two words were 'same' or 'different'. No trial included two words from the same speaker, so there were 54 unique AX pairings. All pairings were played in both orders, for a total of 108 randomly ordered trials. The three categories of stimuli included the fully modal [baɾ] ("M"), breathy C [bʱaɾ] ("C"), and breathy V [bạɾ] ("V"). "SAME" trials paired items within class (MM, CC, VV); "DIFFERENT" trials paired items across class (MC, MV, CV). In the crucial trials are CV ([bʱaɾ] vs. [bạɾ]), which reveal whether the two types of breathy stimuli are reliably distinguished. Overall accuracy by trial-type is presented in Fig. 3.

Figure 3: Mean accuracy in AX discrimination. * = responses significantly different from chance.



Chi-square tests compared the accuracy of each trial type to chance (here: 50%). Listeners correctly identified MM and CC trials as the same, performing significantly above chance ($p < .0001$), and reliably identified these stimulus pairs as different in MC trials ($p < .0001$), but were not above chance in the target CV trials differentiating [bʱaɾ] from [bạɾ] ($p = .1136$). Breathy V [bạɾ] stimuli were problematic in general. In VV trials, listeners identified two breathy vowel stimuli as being the same at just 61.1% accuracy, not significantly above chance (p = .0593). In MV trials ([baɾ] vs. [bạɾ]), their accuracy of 31.9% was significantly *below* chance (p < .0001), meaning participants were reporting them to be the same.

3.3. Identification task

The identification (ID) task sought to determine overlap between categorization of the target words. Unlike the previous tasks, in the ID task categories were defined by the experimenters. Listeners heard an audio stimulus and saw an image simultaneously, and were asked whether the image represented the lexical item in the audio. Like the discrimination task, there were SAME trials, wherein the audio and image matched, and DIFFERENT trials, where they did not. Mean accuracy rates are in Table 2. An asterisk indicates a result that differs significantly from chance (here: 50%).

Table 2: Percentage SAME response in ID task. In shaded cells, audio and picture matched (correct answer: "same"). In unshaded cells, audio and picture differed (correct answer: "different"). Greater accuracy is indicated by high values in shaded boxes and low values in unshaded boxes. Asterisks indicate response rates that differ significantly from chance (50%).

		Audio		
		[baɾ]	[bạɾ]	[bʱaɾ]
Image	[baɾ] 'twelve'	97.5*	70*	5*
	[bạɾ] 'outside'	65	62.5	42.5
	[bʱaɾ] 'burden'	17.5*	22.5*	70*

The trend here is similar to the discrimination task. In SAME trials, listeners accurately identified that the image and audio matched for fully modal [baɾ] 'twelve' and for breathy consonant [bʱaɾ] 'burden', but were not above chance in doing the same for breathy vowel [bạɾ] 'outside'.

In DIFFERENT trials, listeners identified with above-chance accuracy the mismatch between the image for /bʱaɾ/ 'burden' and the audio of both [bạɾ] and [baɾ]. However, listeners did not perform significantly differently from chance when given the image for /bạɾ/ 'outside', regardless of the audio. Most interestingly, the mismatch between the image for /baɾ/ 'twelve' and the audio [bʱaɾ] was identified with above-chance accuracy, but that same image was also identified as a match with the audio [baɾ], with above-chance (in)accuracy. That is, listeners correctly indicated that the audio [bạɾ] did not correspond with the image for /baɾ/ 'twelve' only 30% of the time; the other 70% of the time they reported that the audio and image *matched*. And, with a response rate significantly below chance (p = .0114), this means they were not guessing, rather they were asserting that the breathy vowel audio corresponded with the image of the fully modal word.

4. DISCUSSION

Two results are important to highlight here: (1) the inability of listeners to discriminate between [bạɾ] with a breathy vowel and [bʱaɾ] with a breathy consonant; and (2), the inability of listeners to reliably identify that [bʱaɾ] with a breathy consonant does not correspond with the image for /bạɾ/ 'outside' with a breathy vowel. Both results suggest that breathy consonant and breathy vowel sequences are not reliably differentiated by listeners.

The discrimination task most directly addressed the salience of the difference between any two categories. Presentation of two audio stimuli in immediate succession should cause listeners to compare their acoustic properties without having to categorize them phonologically [15]. Yet, listeners' responses suggest that the acoustic differences between /bʱaɾ/ with a breathy consonant and /bạɾ/ with a breathy vowel are not sufficiently robust: listeners deemed these stimuli "different" at chance.

In the ID task listeners were willing to identify breathy consonant [bʱaɾ] audio as corresponding to images of breathy vowel /bạɾ/ 'outside' but unwilling to do the inverse, i.e. identify breathy vowel [bạɾ] audio stimuli as corresponding to images of breathy consonant /bʱaɾ/ 'burden'. This is likely due to two factors: the robust breathiness associated with the offset of breathy consonants, plus an unexpected ambiguity as to its association. That is to say, listeners reliably identify the phonetic presence of breathiness in [bʱaɾ] audio, but are willing to assign it to either the consonant or the vowel, so it is deemed an acceptable realization of either /bʱaɾ/ or /bạɾ/. This supports both the hypothesis that the two types of breathy stimuli are not well distinguished, and that vocalic breathiness is weakly cued.

Listeners were also at chance when provided with a matched pair (SAME) of breathy vowel audio and image in the ID task, and when presented with two breathy vowel stimuli in the AX task, suggesting that the breathiness associated with vowels is variable in a way that consonant breathiness is not: listeners do not reliably perceive breathiness in the breathy vowel stimulus [bạɾ], and are therefore unwilling to consider such stimuli as realizations of a word that should have robust breathiness, i.e. breathy consonant /bʱaɾ/. The confusion runs in only one direction, though: breathy consonant [bʱaɾ] can be mistaken as a realization of underlying breathy vowel /bạɾ/, but not vice versa. Rather, listeners more consistently accepted breathy vowel [bạɾ] audio stimuli as realizations of an image representing underlying all-modal /baɾ/.

These findings suggest that breathy vowel [bạɾ] is rarely identified as breathy consonant /bʱaɾ/ because vocalic breathiness is so subtle it will more likely pass

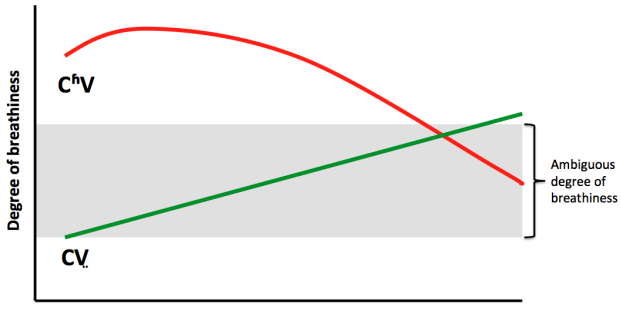
for all-modal /baɾ/ (cf. [14]). The free sort results also indicate an increased probability of overlap between breathy vowel [bạɾ] and all-modal [baɾ], which were put into a single group by some listeners while breathy consonant [bʱaɾ] tended to remain distinct across all response patterns. And in the discrimination task, performance was below chance in trials involving all-modal and breathy vowel stimuli, indicating that listeners reliably consider breathy vowel stimuli and all-modal stimuli to represent the same word. If all-modal [baɾ] can serve as a realization of underlying breathy vowel /bạɾ/, listeners may reliably hear the difference between stimuli of each type yet consider them acceptable variants of the same word.

For the ID task results to be consistent with the hypothesis that /CV̰/ can be realized as [CV] but not vice versa, listeners should identify all-modal audio [baɾ] as a match with an image representing breathy vowel /bạɾ/ 'outside'. The results are mixed, as listeners did not perform significantly different from chance in this specific pair. Interestingly, the inverse pattern did occur, however: the breathy vowel audio [bạɾ] was reliably identified as a match with the image for all-modal /baɾ/ 'twelve'. The weak breathiness in the breathy V audio was not salient enough to prevent listeners from identifying the word as fully modal.

We argue these results can be explained as an effect of inadequate cues to vocalic breathiness. We propose that breathiness functions like other continuous variables that are perceived categorically (e.g. VOT), cued by a suite of continuous variables representing spectral tilt, spectral balance, and noise. If the strength of the acoustic cues for breathy vowels lies near the perceptual threshold between breathiness and modality but those for breathy consonants do not, the breathiness of breathy consonant stimuli (CʱV) should be easily identifiable while that of breathy vowel stimuli (CV̰) should be more ambiguous. Breathy consonants are sufficiently breathy so as to pass for either breathy consonants (CʱV) or vowels (CV̰), while breathy vowels are insufficiently breathy and can thus pass for fully modal sequences (CV). Our proposal is schematized in Fig. 4. The vowel after a breathy consonant (CʱV) is represented with intense breathiness at first before a gradual decrease, and sequences of modal consonant and breathy vowel (CV̰) is represented with more moderate, less dynamic breathiness. The breathiness associated with consonants exceeds the zone of ambiguity; the breathiness of vowels does not.

In the scenario proposed by this explanation, listeners are sensitive to the presence of breathiness, provided that it exceeds the zone of ambiguity. Significant cues to breathiness may be sufficient cause for excluding a stimulus from being identified as modal, but insufficient cause for determining if the breathiness is associated with the C or V. The results of the present study strongly suggest that this interpretation merits further investigation.

Figure 4: Schematization of degree of breathiness across timecourse of vowel.



5. CONCLUSION

This study investigated the perception of sequences involving breathy Cs (CʱV), breathy Vs (CV̰), and no breathiness (CV) by native listeners of Gujarati. Listeners reliably perceive the intense breathiness characteristic of breathy Cs (CʱV), but are unable to determine whether that breathiness is associated with the C or the following V. They do not reliably perceive the subtle breathiness characteristic of breathy Vs (CV̰), often indicating these sequences to be equivalent to fully modal sequences (CV). The overarching trend, then, is that [CʱV] can be interpreted as either /CʱV/ or /CV̰/, while [CV̰] is often indistinguishable from /CV/. While ongoing work will further explore the specifics of these trends, it is evident from this study that there is a problem in differentiating [CʱV] and [CV̰] sequences as well as an overlap in either the categorization or perception of [CV̰] and [CV].

7. REFERENCES

[1] Gordon, M. & Ladefoged, P. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29, 383–406.

[2] Ohala, M. 1983. *Aspects of Hindi phonology* (Vol. 2). Motilal Banarsidass Publishers.

[3] Dixit, R. P. 1989. Glottal gestures in Hindi plosives. *Journal of Phonetics* 17, 213–237.

[4] Khan, SD. 2010. Bengali (Bangladeshi Standard). *Journal of the International Phonetic Association* 40(2), 221–225

[5] Berkson, K. H. 2019. Acoustic correlates of breathy sonorants in Marathi. *Journal of Phonetics* 73, 70–90.

[6] Esposito, C. M. 2010. Variation in contrastive phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association* 40(2), 181–198.

[7] Munro, P., & Lopez, F. H. 1999. *Di'csyonaary X:tèe'n Dìi'zh Sah Sann Lu'uc: San Lucas Quiaviní Zapotec Dictionary*. Chicano Studies Research Center.
[8] Jones, T. E., & Knudson, L. M. 1977. Guelavía Zapotec phonemes. *Studies in Otomanguean phonology* 54, 163–180.
[9] Traill, A. 1985. Phonetic and phonological studies of !Xóõ Bushman. Hamburg: Helmut Buske.
[10] Miller, A. L. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35(1), 56–84.
[11] Esposito, C. M. & Khan, S. D. 2012. Contrastive breathiness across consonants and vowels: a comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association* 42(2), 123–143.
[12] Khan, S. D. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40, 780–795.
[13] Clopper, C. G. 2008. Auditory free classification: Methods and Analysis. *Behavior Research Methods* 40(2), 575–581.
[14] Fischer-Jørgensen, E. 1967. Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics* 28, 71–139.
[15] Key, M. 2012. Phonological and phonetic biases in speech perception. Ph.D. diss. University of Massachusetts, Amherst.