Christina M. Esposito*, Sameer ud Dowla Khan,
Kelly H. Berkson and Max Nelson

Distinguishing breathy consonants and vowels in Gujarati

https://doi.org/10.1515/jsall-2019-2011

Abstract: Across languages, the acoustic and articulatory correlates of breathiness are similar whether they are associated with consonants or with vowels. This raises the question of whether breathy consonants are confusable with breathy vowels in languages in which a phonemically breathy vowel contrasts with a phonemically modal vowel that follows a breathy-aspirated consonant, e. g. Gujarati /ba̤ɾ/ 'outside' vs. /bʱaɾ/ 'burden', respectively. We investigate the perception of a minimal triplet of Gujarati words, with a breathy vowel vs. a breathy consonant vs. an all-modal sequence, via three tasks: free-sort, AX discrimination, and picture-matching identification. Results across the three tasks indicate that breathiness is indeed confusable across the association types. Specifically, while listeners do recognize the stronger breathiness in vowels following breathy consonants, they are not necessarily able to determine whether that breathiness is associated with the vowel or the consonant. Furthermore, they do not reliably recognize the subtler breathiness of breathy vowels, which often indicates that they are the same as or an acceptable realization of an all-modal sequence (/baɾ/ 'twelve'). This suggests a potential perceptual merger in Gujarati, despite previously-reported evidence of a robust three-way contrast in production.

Keywords: breathiness, voice quality, phonation, Gujarati, perception

***Corresponding author: Christina M. Esposito,** Macalester College, 1600 Grand Avenue, St. Paul, MN 55105, USA, E-mail: esposito@macalester.edu
Sameer ud Dowla Khan, Reed College, 3203 SE Woodstock Boulevard, Portland, OR 97202, USA, E-mail: skhan@reed.edu
Kelly H. Berkson, Indiana University Bloomington, 1020 E Kirkwood Avenue, Bloomington, IN 47405, USA, E-mail: kberkson@indiana.edu
Max Nelson, University of Massachusetts Amherst, N408 Integrative Learning Center, 650 North Pleasant Street, Amherst, MA 01003, USA, E-mail: manelson@umass.edu

1 Introduction

Languages tend to contrast phonation[1] types on either consonants, as in Indic languages (e. g. Bengali [Khan 2010], Hindi [Ohala 1983; Dixit 1989], Maithili [Yadav 1984], and Marathi [Berkson 2019]), or on vowels, as in Oto-Manguean languages (e. g. Zapotec [Jones and Knudson 1977; Munro and Lopez 1999; Avelino 2010; Esposito 2010b], Jalapa Mazatec [Kirk et al. 1993; Blankenship 2002; Garellek and Keating 2011], and Trique [DiCanio 2010]), but rarely on both consonants *and* vowels. Only a handful of languages contrast both breathy vowels and breathy-voiced aspirated consonants (Esposito and Khan (to appear)); as far as we know, this list is limited to !Xóõ (Traill 1985; Garellek 2019b), Jul'hoansi (Miller 2007), Wa (Watkins 1999, Watkins 2002), White Hmong (Esposito and Khan 2012), and Gujarati (Esposito and Khan 2012).[2]

The contrast between breathy vowels and breathy-voiced aspirated consonants (henceforth, "breathy consonants") is particularly interesting in that, due to aerodynamic restrictions, breathiness phonologically associated with stop consonants is phonetically realized as a breathy-voiced aspirated release into the following vowel (rather than during the stop closure itself). Thus, both breathy vowels and (prevocalic) breathy-voiced aspirated stops involve breathiness during a vowel. This is demonstrated in Figure 1, a spectrogram showing a minimal triplet contrasting phonation sequences in Gujarati: all-modal બાર [bar] 'twelve' ([CV]), breathy consonant ભાર [bʱar] 'burden' ([CʱV]), and breathy vowel બહાર [ba̤r] 'outside' ([CV̤]). The breathy consonant shows a clear interval of intense noise across the spectrum at the vowel onset that dissipates towards the vowel midpoint, while the breathy vowel shows a more subtle distinction from its all-modal counterpart, with a slight increase in noise and diminished formant strength in the higher frequencies throughout the vowel.

Because stop breathiness is realized in the following vowel, an environment for ambiguity between [CʱV] and [CV̤] sequences is created. The question arises: how do speakers distinguish [CʱV] from [CV̤], if at all? Furthermore, if the

1 Phonation is the production of sound via the vibration of the vocal folds. Cross-linguistically, creaky, modal, and breathy voice are the most common phonation categories, but only the latter two categories are relevant here. Modal phonation is characterized by vocal folds with normal adductive and longitudinal tension. Breathy phonation is produced with minimal adductive tension, weak medial compression, and low longitudinal tension. For a review of phonation, see Laver (1980); Ní Chasaide and Gobl (1997); Gordon and Ladefoged (2001); Garellek (2019a); Esposito and Khan (to appear).

2 Many languages contrast breathy vowels and *voiceless* aspirated consonants, e. g. Jalapa Mazatec (Kirk et al. 1993; Blankenship 2002; Garellek and Keating 2011) and Suai (Abramson et al. 2004), but these two categories differ in voicing.

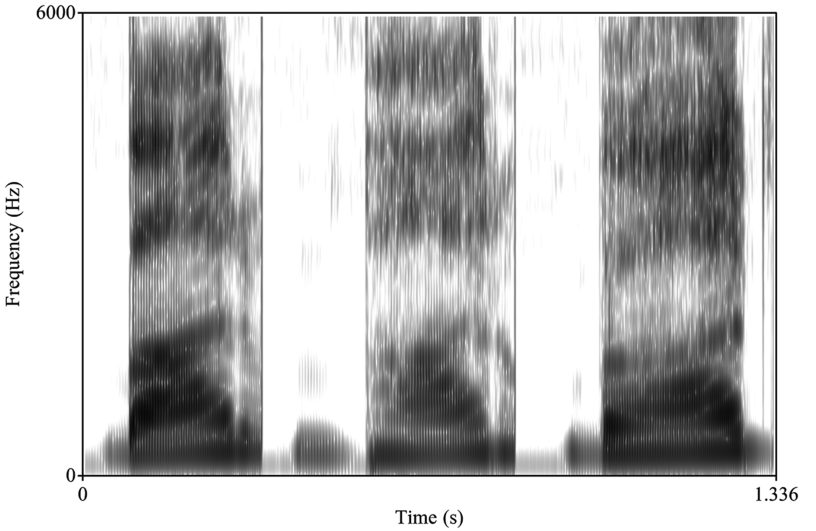

Figure 1: Spectrogram of modal CV in બાર /bar/ 'twelve' (left), breathy consonant C^{ɦ}V in બાર /b^{ɦ}ar/ 'burden' (middle), and breathy vowel C$\underset{\cdot\cdot}{V}$ in બહાર /b$\underset{\cdot}{a}$r/ 'outside' (right).

breathy vowel [C$\underset{\cdot\cdot}{V}$] can appear so similar to the all-modal [CV], how will listeners handle the three-way comparison of [CV], [C$\underset{\cdot}{V}$], and [C^{ɦ}V]? The goal of the current study is to answer these two questions by conducting a three-part perceptual experiment – a free sort task, a discrimination task, and an identification (ID) task – testing speakers of Gujarati.

The paper is divided as follows: the remainder of Section 1 provides background information on Gujarati and on the contrast of breathy vowels vs. breathy-aspirated consonants. We present the current study in Section 2, with general methods and hypothesized results. Detailed methods and results are presented for the free-sort task in Section 3, for the discrimination task in Section 4, and for the ID task in Section 5. Section 6 presents a summary of the results of all three tasks, and Section 7 provides a discussion of these results and concludes the paper.

1.1 Gujarati

Gujarati is spoken primarily in Gujarat, India, by approximately 59 million speakers (Eberhard et al. 2019). Like most Indic (i. e. Indo-Aryan branch of Indo-European) languages, Gujarati boasts a four-way contrast between

voiceless, voiceless aspirated, voiced, and breathy obstruents (Table 1), but unlike other Indic languages it also contrasts modal vs. breathy vowels (Table 2).

Table 1: Four-way phonation contrast in Gujarati stops.

	Orthography	IPA	Gloss
Voiceless unaspirated	કાલ	kal	'yesterday/tomorrow'
Voiceless aspirated	ખાળ	kʰaɭ	'drain'
Voiced unaspirated	ગાળ	gaɭ	'curse word', 'filter!'
Breathy-voiced aspirated	ઘાલ	gʱal	'penetrate!'

Table 2: Two-way phonation contrast in Gujarati vowels.

	Orthography	IPA	Gloss
Modal	કાન	kan	'ear'
Breathy	કહાન	kn̤an	'Krishna'

The most extensive characterizations of the oral vowel inventory of Gujarati include eight modal vowels [i e ɛ a ɔ o u ə] and eight phonemically breathy vowels [i̤ e̤ ɛ̤ a̤ ɔ̤ o̤ ṳ ə̤].[3] However, the realization of the breathy vowels can vary across dialect (Modi 1986) and register (Mistry 1997), which can greatly reduce the number of vowels seen on the surface. For example, breathy vowels commonly alternate with [VɦV] sequences, reflecting the historical source (and current orthographic representation) for most breathy vowels (Dave 1967); as a result, speakers who are reading carefully or trying to speak in a more "educated" register may produce a [VɦV] sequence rather than a single breathy vowel [V̤] (Mistry 1997; Cardona and Suthar 2003; Khan 2012). When the first vowel in this sequence is historically a schwa (i. e. [əɦV]), the usual pronunciation in modern Gujarati is almost exclusively a single breathy vowel in connected speech (i. e. [V̤], Dave 1967), and only arises as a disyllabic sequence (i. e. [əɦV]) in reading, presumably as a spelling pronunciation. For example બહેન [bɛ̤n] 'sister' has a disyllabic variant [bəɦen] in careful or read speech. When the first vowel is anything other than a schwa, there can be much more variation

3 The terminology used in reference to breathy vowels in Gujarati varies in the literature, with the term "murmur" being used notably by early studies e. g. Pandit (1957); Dave (1967); Fischer-Jørgensen (1967); Nair (1979); and Modi (1986). However, for the sake of consistency and keeping with the terminology most prevalent in more recent studies the term "breathy" is used throughout the present work.

(i. e. [VfiV ~ Vfi ~ V̤]); this is the case for words like રાહબર [rafiəbər ~ ɾafibər ~ rə̤bər] 'guide'. Breathy vowels also arise in very casual registers and/or very fast speech rates before breathy consonants, replacing the post-breathiness of the consonant such that [VC^{fi}] may become [V̤C] (Dave 1967; Mistry 1997; Cardona and Suthar 2003).

To further complicate the contrastive relation between breathy and modal vowels, many if not most varieties of Gujarati show some degree of merger between voice quality categories, often interacting with vowel quality, as Cardona and Suthar (2003) describe in detail. For example, while high and high-mid /i̤ e̤ o̤ ṳ/ are frequent in the lexicon, their breathy counterparts /i̤ e̤ o̤ ṳ/ are extremely infrequent except in the most casual registers, where they arise almost exclusively due to interactions of breathy stops and vowels in fast speech rates. Furthermore, while low-mid vowels /ɛ/ and /ɔ/ do show a breathy and modal contrast (e. g. મેલ /mɛl/ 'dirty' vs. મહેલ /mɛ̤l/ 'palace'), breathy /ɛ̤/ and /ɔ̤/ are far more frequent in the lexicon than modal /ɛ/ and /ɔ/, and for some speakers, low-mid modal /ɛ/ and /ɔ/ merge completely with their high-mid counterparts /e/ and /o/, respectively, leaving asymmetrical sets such as /mel/ 'dirty' vs. /mɛ̤l/ 'palace' for such speakers (Dave 1967). And Pandit (1957); Fischer-Jørgensen (1967); Dave (1967); and Nair (1979) all report that there exists some amount of unconditioned variation between breathy and modal vowels in "the speech of many educated Gujarati speakers" (Pandit 1957: 170), but that it is not bidirectional: a modal vowel [V] can serve as an acceptable production of a breathy vowel [V̤] (e. g. /bə̤ɾ/ 'outside' can be pronounced [baɾ]), but the inverse is not true (e. g. /baɾ/ 'twelve' cannot be pronounced *[bə̤ɾ]).

1.2 Gujarati breathiness: Production

Almost all phonetic accounts of breathiness in Gujarati focus on breathy vowels, rather than breathy consonants; brief acoustic descriptions of voice onset time (VOT) are largely all that has been documented for the breathy consonants of Gujarati (e. g. Rami et al. 1999). This bias may be rooted in the fact that while breathy vowels are rare in Indic languages, making Gujarati locally unique, breathy consonants are extremely common in the region.

An early impressionistic account of breathy vowels in Gujarati characterized them as involving voiced breath at a low pitch (Pandit 1957). Later, Dave (1967) and Fischer-Jørgensen (1967) applied more quantitative approaches, finding that Gujarati breathy vowels were characterized by greater airflow than modal ones, but that there was no significant difference between the intensity, formant frequencies, or the formant bandwidths of modal and breathy vowels. The

most prominent spectral property of breathy vowels was a greater amplitude of the fundamental frequency (i. e. amplitude of the first harmonic, H1). Fischer-Jørgensen (1967) also described breathy consonants in Gujarati as similar to a breathy vowel, with the main difference being the degree of noise.

More recently, in an acoustic and electroglottographic (EGG) study on phonation in Gujarati, Khan (2012) found that several harmonic-based spectral measures[4] (including H1*–H2*), noise measures, and closed quotient (an EGG measure) were all reliable indicators of breathy phonation, consistently distinguishing breathy and modal vowels. Furthermore, phonation was localized, especially along noise measures, to the vowel midpoints. These results indicate that phonation is a multi-dimensional space, and changes as a function of time. Nara (2017) extended this investigation to both native and heritage speakers of Gujarati in Canada, finding that while heritage speakers have a phonetically less extreme contrast than native speakers, both groups produce breathy vowels distinctly from modal vowels.

Using the recordings from Khan (2012), Keating et al. (n.d.) compared the breathy vowels of Gujarati to a database of breathy vowels, lax vowels, and vowels of other voice qualities in an array of languages, finding that the breathy vowels of Gujarati are acoustically more similar to the so-called lax vowels of the Yi languages of southern China than they are to the (acoustically more extreme) breathy vowels of White Hmong and Valley Zapotec. This suggests that Gujarati breathy vowels are acoustically less breathy than those of other languages. (Keating et al. did not look at consonant breathiness, however.)

1.3 Comparison of breathy vowels and breathy consonants

Previous research suggests that breathy vowels and breathy consonants share a number of articulatory features. Both types of sounds involve vocal folds that vibrate with little contact, along with high rates of airflow and a moderately open glottis (Ladefoged and Maddieson 1996; Dixit 1989; Gordon and Ladefoged 2001). And, for both consonants and vowels, breathy voice is associated with

4 One of the most common spectral measures is the difference between the amplitude of the first and second harmonics. See Gordon and Ladefoged (2001) for an overview on measuring phonation from an audio signal. An asterisk indicates that the harmonic amplitude (e. g. H1*) has been corrected for influences of formant frequencies and bandwidths. Correction algorithms for harmonics have only recently been applied to linguistic voice studies, and therefore may not appear in research before 2005.

increases in spectral balance, spectral slope, and noise (Berkson 2013, Berkson 2019; Dutta 2007; Esposito 2010a; Khan 2012).

Despite the similarities, a comparative study of Gujarati and White Hmong (Esposito and Khan 2012) demonstrated that breathiness associated with consonants can be distinguished acoustically and articulatorily from breathiness associated with vowels. In both languages, timing and degree distinguished consonantal breathiness from vocalic breathiness: breathy consonants were characterized by a short period of intense breathiness at the onset of the vowel followed by decreasing breathiness, while breathy vowels showed stable (Gujarati) or increasing (White Hmong) breathiness across the timecourse of the vowel.

1.4 Gujarati breathiness: Perception

While there have not been previous studies of the perception of breathy consonants in Gujarati, a number of studies have explored the perception of breathy vowels by Gujarati listeners. Fisher-Jørgensen (1967) observed that the differences between breathy and modal vowels in Gujarati were very small, even to a trained ear, and decided to see if Gujaratis could distinguish breathy vowels from modal ones in their own language. Results showed that some listeners were able to successfully identify breathy vowels, especially in cases where the spectrum was dominated by H1 amplitude. Other cues, such as duration and f0, had little importance in the identification of breathy vowels. Interestingly, listeners had difficulty in this task: even when they heard clearly produced natural examples of minimal pairs, they had trouble identifying breathy vowels, relative to identifying modal vowels. This was true even though the listeners themselves served as talkers in another task, producing the same stimuli and exhibiting consistent acoustic distinctions in their own pronunciation.

Bickley (1982) also examined the perceptual correlates of breathiness by obtaining judgments from Gujarati listeners. Listeners identified words containing synthesized breathy vowels, in which H1 amplitude and the degree of aspiration noise varied. Results showed that vowels with the highest H1 amplitude were consistently judged to be breathy.

Esposito (2006, 2010a) investigated the perception of phonation by listeners of Gujarati, Spanish (which has neither phonemic nor allophonic breathiness), and English (which has allophonic breathiness). Results showed that Gujarati listeners' judgments were strongly correlated with H1*–H2* and that their judgments were more consistent than the other two listener groups. The listeners' experience as Gujarati speakers helped them focus directly on H1*–H2*, the

single most important acoustic cue to breathiness in their language, whereas speakers of other languages grasped at a range of cues.

In another cross-linguistic perception study of breathiness, Kreiman et al. (2010) calculated the just-noticeable difference (JND) in H1*–H2* between synthesized vowels for listeners of Gujarati and three languages that do not have phonemic voice quality contrasts: Mandarin, Thai, and English. The authors found that, as expected, Gujarati-speaking listeners had the smallest JNDs of all the four groups of listeners, meaning they were sensitive to the smallest differences in H1*–H2*. Interestingly, this effect held regardless of whether the listener self-identified as a Gujarati-dominant speaker who had spent little time outside India or a heritage speaker of Gujarati who had spent most or all of their life in the US; even passive fluency in Gujarati appears to be sufficient to gain heightened sensitivity to subtle distinctions in breathiness. Note, however, that we cannot be sure from these results if Gujarati listeners would also be able to leverage their heightened sensitivity in order to distinguish a three-way contrast in breathiness; to our knowledge, such a contrast has not previously been subject to a perception task.

2 Current study

The present study is a preliminary account of Gujarati-speaking listeners' ability to differentiate and categorize [CV], [C$^{\text{fi}}$V], and [C̤V̤] sequences produced by Gujarati-speaking talkers. Given previous findings in production, we predict a three-way distinction in distinguishing and categorizing [CV], [C$^{\text{fi}}$V], and [C̤V̤] sequences, in line with native productions. And, given findings in perception, we predict that native and heritage speakers of Gujarati should have no trouble identifying these sequences, even when they are subtle. However, because no previous study has examined the perception of all three categories in Gujarati, we may alternatively predict that even native speakers will have trouble with a three-way distinction, and we are especially interested in how the breathy vowels [C̤V̤] will be interpreted given their similarities with both breathy consonants [C$^{\text{fi}}$V] and all-modal [CV] sequences, and their complicated status in Gujarati phonology. Due to these conflicting hypothesized results, we take as our first step, an exploratory perception experiment, using a small sample of Gujarati speakers, and a narrow set of stimuli. The results of this more constrained experiment will still yield important results: are Gujarati speakers able to distinguish these three stimulus types? The results presented here will lay the groundwork for future research discussed in the conclusion.

2.1 Methods

The experiment was divided into three ordered sections; a free sort task, a discrimination task, and an ID task. Tasks were ordered – rather than randomized across participants – because the design of the ID task imposed three experimenter-defined categories on the participants, so it was crucial to administer it last.

2.1.1 Stimuli and talkers

Stimuli consisted of recordings of a single well-known minimal triplet shown in Table 3. The choice to focus on these three words was based on several considerations. First, this can simplify a complicated set of tasks for the participants. Secondly, all three of the words in this set are expected to be highly familiar to all speakers, and could be visualized in an image (required for the third task). Thirdly, using only the low vowel [a] avoids the potential confounds with vowel height and phonation seen in the low-mid, high-mid, and high vowels (Cardona and Suthar 2003). It has in fact been argued that the most robust cues for breathiness appear in low vowels; previous research has often focused solely on low-vowel contexts (e. g. Dutta 2007). At present there exists no literature addressing the ability of speakers to distinguish [C$^{\text{ɦ}}$V] and [CV̤] sequences, so this more restricted perceptual experiment can lay the groundwork for more in-depth future studies. And, finally, if listeners are not able to distinguish the vowels even in a minimal set, where such differences are crucial, then they will likely not be able to perceive this difference in the language at large.

Table 3: Minimal triplet of breathiness in Gujarati, used for all stimuli.

	Orthography	IPA	Gloss
All modal (no breathiness)	ભાર	bar	'twelve'
Breathy vowel	ભઆર	ba̤r	'outside'
Breathy consonant	ભ્હાર	b$^{\text{ɦ}}$ar	'burden'

All stimuli were taken from Khan (2012); in the original recordings, ten (7F, 3M) native speakers of Gujarati, all recent immigrants to the US, were asked to produce target words utterance-initially, in a self-generated carrier sentence (see Khan 2012 for a fuller description and justification of the procedure). The

stimuli used in the current study were extracted from those sentences, zero-crossed to maximize the naturalness of each token, and judged acceptable by two trained phoneticians.

Because studies have suggested that the cues for breathiness may not be consistent across gender (Berkson 2013; Esposito 2010a), all stimuli were taken from four female talkers. To further control for the effects of sociolinguistic variation by age and region, all talkers were between 20–30 years of age and raised in Mumbai.[5] All four described their speech as being typical of Mumbai Gujarati, with little influence of other varieties.

Two tokens of each word were chosen deliberately from each talker so as to minimize durational differences that may occur between talkers. However, minor vowel durational differences were unavoidable, as we preferred to leave the stimuli as natural as possible. Average duration ranged from 200–250 ms Altogether there were 4 talkers x 3 words x 2 repetitions = 24 stimuli.

H1*–H2* was measured automatically at five points across the duration of the vowel using VoiceSauce (Iseli et al. 2007; Shue et al. 2011). During the first three timepoints, H1*–H2* was higher for vowels after breathy consonants [C$^{\text{ɦ}}$**V**] than for breathy vowels [C**V̤**], while modal vowels [C**V**] had the lowest H1*–H2* value. By timepoint four, vowels in both breathy sequences [C**V̤**], [C$^{\text{ɦ}}$**V**] had more similar H1*–H2* values and both remained higher than modal vowels. In terms of change over the course of the vowel, vowels in all-modal sequences had a low H1*–H2* value across all five timepoints, while vowels following breathy consonants [C$^{\text{ɦ}}$**V**] reached their highest H1*–H2* values at timepoint two; breathy vowels [C**V̤**] were similar to vowels after breathy consonants (C$^{\text{ɦ}}$**V**) but exhibited their highest H1*–H2* values at timepoints four and five. This pattern is illustrated in Figure 2.

2.1.2 Listeners

Six listeners (2F, 4 M) were recruited from Indiana University; one was a fluent heritage speaker born in the US, while the remaining five were native to India. Our choice to focus on speakers in the US was borne both out of convenience and out of the need for consistency across studies: because virtually all recent production and perception work on Gujarati has been based on data from multilingual Gujarati speakers living in North America, we wanted to draw from a similar population for the current study. Also recall that Nara (2017)

5 While Mumbai is outside Gujarat state, it has a large and well-established community of over 1.4 million Gujarati speakers (Blank 2002; Census of India 2011).

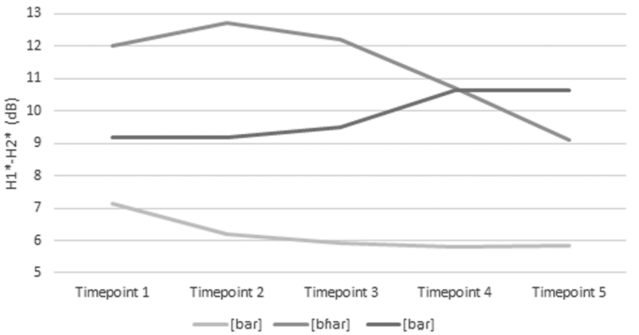

Figure 2: Average H1*–H2* (dB) of the 24 stimuli used in the three perception tasks. The lowest line represents the all-modal [bar], the slightly rising middle line represents the breathy vowel [b̤ar], and the high-falling line represents the dynamic realization of the breathy consonant [bʱar].

found a breathy vs. modal vowel contrasts among both native and heritage speakers of Gujarati in Canada, and that in Kreiman et al.'s (2010) perception study, Gujarati listeners were sensitive to subtle distinctions in breathiness, regardless of whether they were a Gujarati-dominant speaker who had spent little time outside India or a heritage speaker who had spent most or all of their life in the US. As is common in this speech community, all listeners reported fluency in Gujarati as well as other languages and were asked to self-report language dominance. Listener information, including gender, age, birthplace, and language dominance, is presented Table 4.

Table 4: Listener demographics. Note that even listeners who reported greater dominance of other languages self-identified as fluent speakers of Gujarati.

Listener	Gender	Age	Birthplace	Native languages spoken, in decreasing order of dominance
1	M	23	Ahmedabad, Gujarat, India	Gujarati, Hindi, English
2	M	30	Vadodara, Gujarat, India	English, Hindi, Gujarati
3	M	23	Mumbai, Maharashtra, India	Urdu, Gujarati, English, Marathi
4	M	26	Ahmedabad, Gujarat, India	Gujarati, Hindi, English
5	F	52	Bardoli, Gujarat, India	Gujarati, English, Hindi
6	F	19	Fostoria, Ohio, USA	English, Gujarati

Because all listeners were fluent English speakers, and accustomed to using English in academic contexts, directions for all tasks were conveyed in English.

3 Free-sort task

We began the experiment with a free-sort task, in which subjects sort a set of items into groups according to self-chosen criteria. Free-sort tasks have been used in previous studies exploring the perception of voice quality differences (Granqvist and Eng 2003), including studies with Gujarati listeners (Esposito 2006, Esposito 2010a). Essentially, a free-sort task is like an ID task, differing only in the nature of the response. It is an advantageous procedure in that it avoids introducing any experimenter-imposed categories on the stimuli.

3.1 Procedure

The free-sort task investigated whether listeners independently proposed three target categories ([baɾ], [bʱaɾ], and [baɾ]) when presented with a screen containing 24 numbered icons (one for each token) arranged randomly in columns (Figure 3, left) and asked to categorize them by dragging them to the right side of the screen and placing them in groups (see a sample outcome in Figure 3, right). Icons corresponded randomly to one of the 24 audio stimuli, and played when clicked. Listeners were able to both play the icon and move it around the screen multiple times. They were given no time limit, but generally they finished the task in 15–20 minutes.

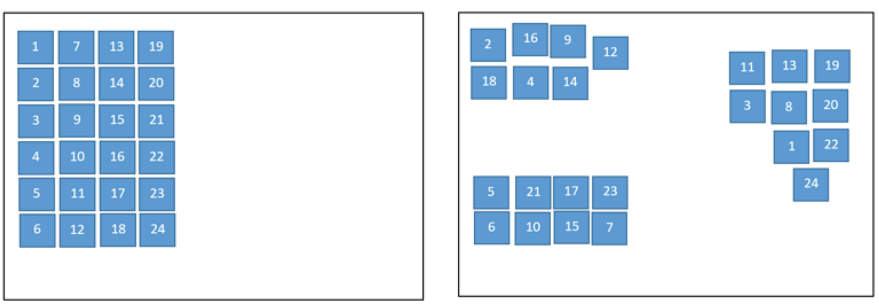

Figure 3: Free-sort task setup (left) and one sample outcome (right).

To avoid experimenter-imposed biases in sorting, participants had absolute freedom over how to categorize the items as well as how many categories to propose. This opened up the possibility that they might categorize the items based on non-target criteria, for example by talker identity. This risk was deemed acceptable as there is potentially crucial information to be gained from participants who do not create three categories. Based on previous observations by Pandit (1957); Dave (1967); Fischer-Jørgensen (1967); and Nair (1979), a listener might group modal vowels into the same space as breathy ones, since the former can pass as acceptable versions of the latter. Thus, forcing participants to create three distinct groups could prevent them from grouping together modal and breathy vowels, even if their intuition would be to place them in the same group.

3.2 Results

Listeners used different approaches to categorization, thus a purely descriptive report of the outcomes is most informative here. Individual results are displayed in Figures 4(a)–4(f), and a summary of the results are presented in Table 5. In Figures 4(a)–4(f), black and white icons represent modal stimuli, light gray icons represent breathy consonant stimuli, and dark gray icons represent breathy vowel stimuli.[6] Numbers on the icons correspond to stimulus number, and the arrangement of icons reflects how the listeners grouped the stimuli.

Listeners 1 and 2 attempted to pair all stimuli by both word and talker, resulting in 12 groups of two stimuli each (Figure 4(a)–4(b)). (When listeners grouped stimuli within a talker, the boxes are displayed next to each other [Figure 4(a)]. When listeners grouped stimuli across talkers, a line is drawn to connect the two boxes [Figure 4(b)]. Talker information is also provided for each row.) This suggests that Listeners 1 and 2 were able to leverage talker-specific acoustic information. Listener 1 was highly accurate in grouping stimuli this way, while Listener 2 was slightly less so.

Listeners 3 and 4 formed two large groups of stimuli each (Figures 4(c)–4(d)): respectively, one of the large groups represents a well-defined breathy consonant category while the other was a grouping of modal and breathy vowels together. This suggests overlap in the modal and breathy vowel categories, as suggested by the literature.

6 In the task itself, all icons were identical in color, and groupings of icons were distributed throughout the screen following the listener's choice, but in Figure 4 the icons have been rearranged and re-colored for visual clarity.

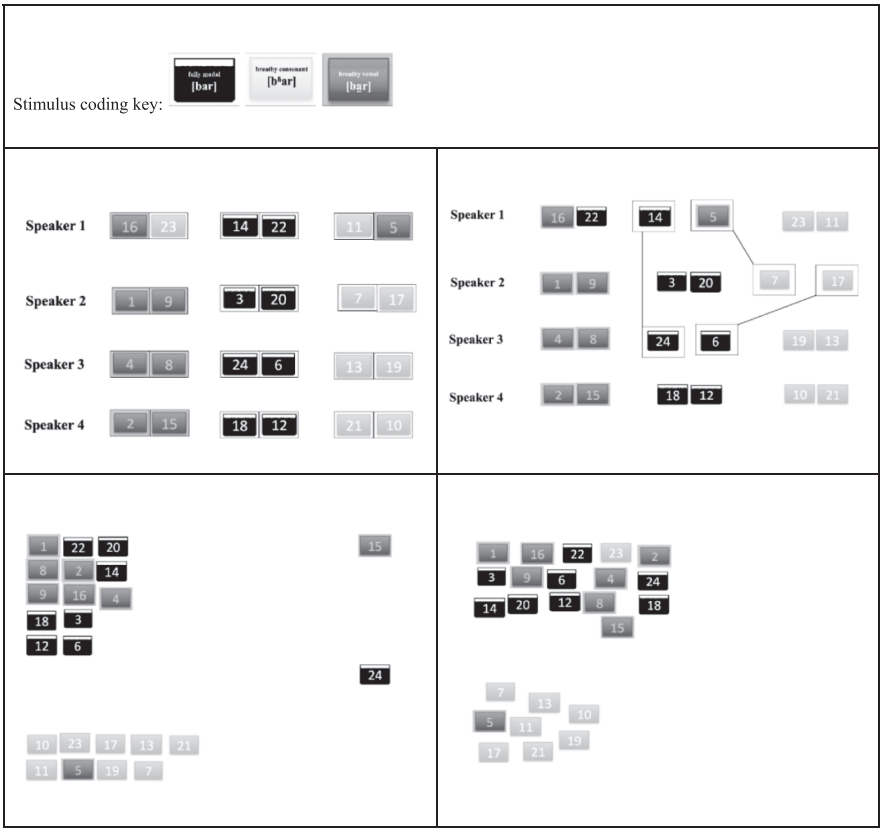

Figure 4: Coded free-sort task results. (a) Listener 1 grouped stimuli by both stimulus type and talker with near-perfect accuracy. The two inconsistent pairs of stimuli (16 & 23, 11 & 5) each group together a breathy vowel and breathy consonant. (a), (b) Listener 2 largely grouped by both stimulus type and talker, but also created groups spanning talker (14 & 24) and stimulus type (16 & 22) and groups inconsistent in both dimensions (6 & 17, 5 & 7), (c) Listener 3 created two primary groups: a well-defined [bʰar] category with one incongruous item (stimulus 5) which is problematic across all participants. The other category combines [bar̤] and [bar]. Two stimuli were set aside. No [bʰar] stimulus is rejected or placed in an inconsistent group, (d) Listener 4, like Listener 3, created two clear groups, one representing [bʰar] and the other combining [bar̤] and [bar]. Again the only incongruent item in the [bʰar] group is stimulus 5. The one [bʰar] stimulus placed in the [bar̤] + [bar] group, stimulus 23, was also miscategorized by Listener 1, (e) Listener 5 created three groups that may be loosely based on the three stimulus types. The horizontal group at the has the greatest concentration of modal stimuli while the lower two groups seem to represent [bʰar] (left) and [bar̤] (right), but all groups are mixed, (f) The groups formed by Listener 6 do not seem to be regular. Some pairs, such the one in the lower-left corner (stimuli 10 & 21), are accurately paired stimulus types of the same talker, while other groups such as that in the top left corner of the screen show little discernible organization.

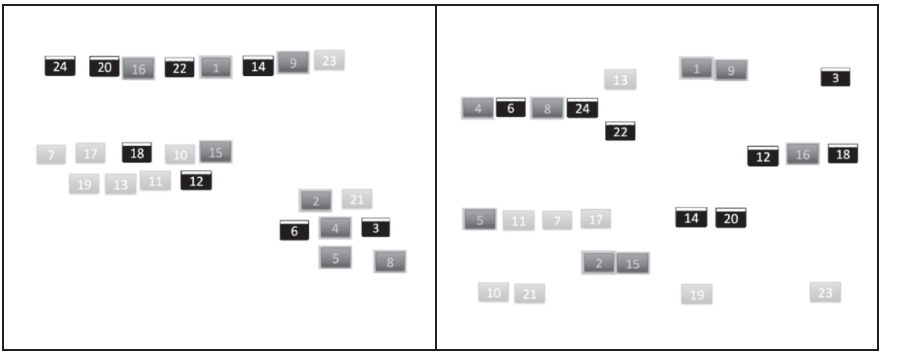

Figure 4: (*continued*).

Table 5: Summary of free-sort task results by listener.

Listener	Interpretation of grouping
1	12 groups: based on word & talker
2	12 groups: based on word & talker
3	2 groups: breathy consonants vs. modal vowels + breathy vowels
4	2 groups: breathy consonants vs. modal vowels + breathy vowels
5	3 groups: breathy consonants vs. modal vowels vs. breathy vowels (roughly)
6	uninterpretable

Listener 5 (Figure 4(e)) created three groups which may have been intended to represent the three categories of stimuli: each consisted of a slight majority of one type of stimuli, but all groups were mixed and contained at least one member of each of the three stimuli types. Even here, however, the most consistently grouped stimuli were breathy consonants, perhaps indicating that these are the least confusable type of stimuli.

The responses of Listener 6 (Figure 4(f)) were not interpretable by any of the authors.

The responses to the free sort task can be divided into three categories: those by listeners who paired by stimulus type (i. e. word) and talker (Listeners 1 & 2), those by listeners who created two distinct groups, /b$^{\text{ɦ}}$ar/ vs. /bar, b̤ar/ (Listeners 3 & 4), and those by listeners who followed other patterns (Listeners 5 & 6). The accuracy with which Listeners 1 and 2 were able to match types of tokens from the same talker suggests that they may be able to accurately differentiate all three types of stimuli from the acoustic information alone, while the results from Listeners 3 and 4 suggest a potential overlap in the modal and breathy vowel categories.

Interestingly, language dominance and place of birth seems to play little role in the results. For example, Listener 1 self-reported dominance in Gujarati, while Listener 2 self-reported dominance of English and Hindi over Gujarati, and yet both employed a very similar strategy. The same is true for Listeners 3 (Urdu-dominant) and 4 (Gujarati-dominant). Inversely, the three listeners born in India who reported dominance in Gujarati (Listeners 1, 4, & 5) showed three different patterns of grouping. And even the two listeners from the same city (Listeners 1 & 4, from Ahmedabad) used very different strategies.

Notably, we do recognize that Listener 6, both a heritage speaker and the only participant born in the US, had uninterpretable results. This is unexpected given Nara's (2017) and Kreiman et al.'s (2010) findings on heritage speakers in North America.

4 Discrimination task

The discrimination task aimed to determine the accuracy with which participants can distinguish pairs of target words. In a sense, it is the task that most directly addresses the issue of perceiving the difference between [CV], [C$^{\text{ɦ}}$V], and [C̰V̰] sequences, because, while in other perceptual tasks the participant may first categorize stimuli and then compare categories rather than the stimuli themselves, a discrimination task encourages a direct comparison of stimuli (Key 2012).

4.1 Procedure

Items were presented in an AX (same–different) task. In order to constrain the total duration of the experiment, only one repetition of each stimulus type from each talker was used in the AX task for a total of 12 stimuli (3 items × 4 talkers). Participants heard two of the 12 stimuli in succession, separated by a 300 ms interstimulus interval, and indicated whether the two were the same or different by pressing a designated computer key.

Pairs played only once, and could not be repeated. Participants had 1000 ms to indicate their response; all responded within the given timeframe. The task took approximately 8 min to complete. Tokens from the same talker were never paired, but otherwise every item was paired with every other item. There were six different types of trials; three in which the stimuli were of the same category (e. g. [baɾ]–[baɾ], [b$^{\text{ɦ}}$aɾ]–[b$^{\text{ɦ}}$aɾ], [ba̰ɾ]–[ba̰ɾ] "same" trials) and three in which the

stimuli were of different categories (e. g. [baɾ]–[bʱaɾ], [bʱaɾ]–[baɾ], and [baɾ]–[baɾ] "different" trials).

4.2 Results

Overall accuracy by trial-type is presented in Figure 5. To confirm that a contrast is perceptually salient, listeners must discriminate stimuli at a rate significantly above chance (in a task with two possible answers, this is 0.5). To confirm that two members of the same word class are perceived as the same, listeners must discriminate stimuli at a rate significantly *below* chance, i. e. demonstrating that they reliably identify the two as not different. Chi-squared tests compared the accuracy of each trial type to chance.

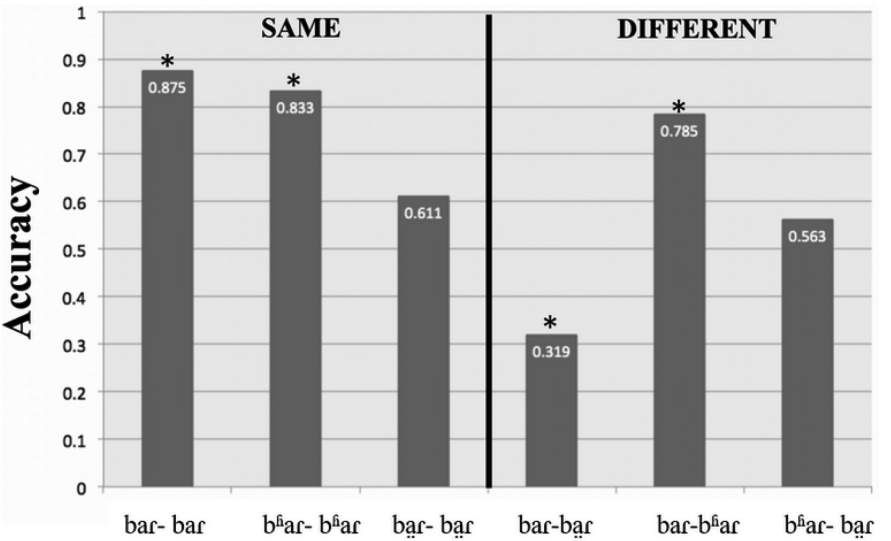

Figure 5: Mean accuracy in the AX discrimination task, across trial type. The asterisk * marks responses that were statistically significantly above or below chance.

In [baɾ]–[baɾ] and [bʱaɾ]–[bʱaɾ] "same" trials, participants had an accuracy rate of 87.5% and 83.3% respectively, performing significantly above chance ($p < 0.0001$). And, listeners were also able to reliably differentiate these two types of tokens as evidenced by their average accuracy (78.5%) in [baɾ]–[bʱaɾ] "different" trials ($p < 0.0001$).

However, in [bạɾ]–[bạɾ] "same" trials listeners did not perform significantly different from chance (p = 0.06), and correctly identified two breathy vowel stimuli as being "the same" with just 61.1% accuracy. In fact, discriminating breathy vowel stimuli was problematic for listeners across the board. Listeners performed poorly on the [baɾ]–[bạɾ] "different" trials, with an average accuracy of 31.9%, significantly below chance ($p < 0.0001$). These two results can be interpreted in the following way: listeners were not just inaccurate in correctly identifying [baɾ] and [bạɾ] as different, they were inaccurate because they actively considered them to be the same. This may suggest an insufficient level of breathiness perceived on breathy vowel tokens.

Furthermore, in [b^{ɦ}aɾ]–[bạɾ] "different" trials, the average accuracy of 56.3% is not significantly above or below chance ($p = 0.11$), corroborating the hypothesis that while listeners do not consistently consider these "the same", they also are unable to reliably tell the two types of breathy sequences apart.

In sum, listeners showed little evidence of distinguishing [baɾ] from [bạɾ]; in trials that paired these stimuli, listeners reliably responded with the *incorrect* answer, indicating that these two sequences were "the same". Listeners also performed at chance when presented with trials that paired [b^{ɦ}aɾ] and [bạɾ] stimuli: they could not accurately distinguish between these two types of breathy voice. These results suggest that while listeners were unsure about the relation between [bạɾ] and [b^{ɦ}aɾ], they were more certain that [bạɾ] and [baɾ] were "the same".

5 Picture-matching ID task

The picture-matching ID task sought to determine overlap between categorization of the target words. Unlike the previous two tasks, the ID task allowed participants to determine whether a stimulus was an acceptable member of an experimenter-defined category. Consider the very real possibility that a native speaker might find [baɾ] to be an acceptable realization of /bạɾ/ 'outside'; even such a speaker might choose not to group them together in the free-sort task where they can play the tokens repeatedly and deliberate about how to group them, and they may respond to subtle acoustic differences between [baɾ] and [bạɾ] in the discrimination task. But in an ID task, they might still indicate that [baɾ] *can* correspond to the meaning of /bạɾ/ 'outside'.

5.1 Procedure

Listeners were presented with one of three images (Figure 6) representing the three target words which appeared immediately after presentation of one of the audio stimuli, and asked to indicate if the audio and the image where the "same" or "different". (A native speaker of Gujarati assisted in selecting the picture representations of the target words.) As there were three different types of audio stimuli and three images representing the meaning of the stimuli, there were nine different types of trials. Pictures were used rather than orthography in order to prevent any potential confounding effects that may arise from the way breathy vowels are represented in Gujarati orthography. This task took approximately 5 min to complete.

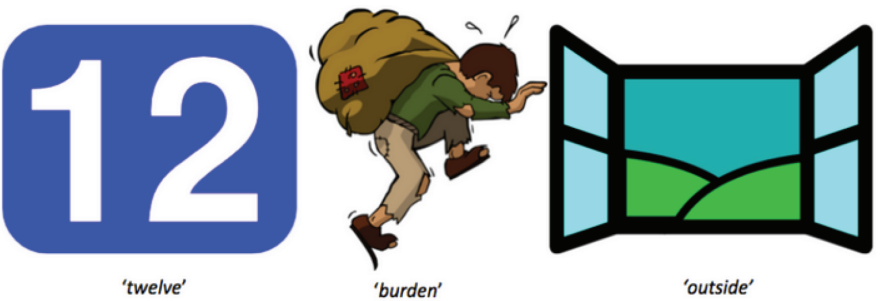

Figure 6: Images representing one of the three target words: બાર /bar/ 'twelve', ભાર /bʱar/ 'burden', and બહાર /bə̰r/ 'outside'.

5.2 Results

Average rates of "same" responses in the ID task appear in Table 6.

For "same" trials, listeners accurately identified that the picture and audio stimulus matched in trials with [bar] 'twelve' (average accuracy: 97.5%) and [bʱar] 'burden' (average accuracy: 70%). However, they did not perform significantly differently from chance in identifying that the audio [bə̰r] matched the image for /bə̰r/ 'outside' (average accuracy = 62.5%), indicating that there was difficulty with the perceptual salience of breathy vowels as a category.

For "different" trials, listeners correctly identified mismatches between [bʱar] and [bar] with above-chance accuracy, in both permutations of image vs. audio, suggesting that these two categories are robustly distinct to our listeners. In trials in which the audio [bʱar] is paired with the image for /bə̰r/

Table 6: Percentage "same" response in ID task. Shaded cells are matches, i. e. those where the correct answer was "same", while unshaded cells are mismatches, i. e. those where the correct answer was "different". Greater accuracy is indicated by higher values in shaded boxes and lower values in unshaded boxes. Asterisks * indicate response rates that differed significantly from chance (50%), and the exclamation mark ! represents the cell in which results significantly differed from chance in the unexpected direction (i. e. a mismatch heard as a match).

		Audio		
		[bar]	[ba̰r]	[b$^{\text{ɦ}}$ar]
Image	/bar/ 'twelve'	97.5*	70.0* !	5.0*
	/ba̰r/ 'outside'	65.0	62.5	42.5
	/b$^{\text{ɦ}}$ar/ 'burden'	17.5*	22.5*	70.0*

'outside', however, listeners were not significantly accurate at identifying that the word and picture did not correspond (average accuracy: 57.5%). In the inverse situation, when participants were presented with the audio [ba̰r] and the image /b$^{\text{ɦ}}$ar/ 'burden', they were able to identify the mismatch at a rate significantly above chance (average accuracy: 77.5%).

Listeners were not able to accurately identify that the audio [bar] was *not* a match with the image of /ba̰r/ 'outside'; they performed at a rate not significantly different from chance (average accuracy = 35%). Furthermore, participants correctly indicated that the [ba̰r] audio did *not* correspond with the image for modal /bar/ 'twelve' only 30% of the time; instead, they consistently reported that breathy vowel audio [ba̰r] was a *match* with the image of the modal word /bar/ 'twelve' at a rate significantly different from chance (p = 0.01; marked with ! in Table 6). As a whole, these results suggest that modal and breathy vowels have significant perceptual overlap, although not to the point of complete merger, given the directional bias. This is especially noteworthy in that we had predicted some potential overlap here in line with previously reported descriptions of variability in breathy vowels.

These results indicate listeners were occasionally willing to identify audio [C$^{\text{ɦ}}$V] as a realization of /CV̰/, but not willing to identify audio [CV̰] as a realization of /C$^{\text{ɦ}}$V/, which could suggest a fine-tuned ability to distinguish two types of breathiness. However, another explanation becomes apparent when observing that the mismatch between the image of /bar/ 'twelve' and the audio of [ba̰r] 'outside' was *not* correctly rejected; on the contrary, listeners considered the breathy vowel audio to be an acceptable realization of the image

representing the fully modal word, indicating that they did not perceive the breathiness on [bạr] 'outside'. At significantly below chance, this suggests that they were not guessing: rather, they actively indicated that the breathy vowel audio [bạr] corresponded with the representative image of the fully modal word /bar/ 'twelve' with more confidence than when the image was of the breathy vowel word /bạr/ 'outside'.

6 Summary of all three tasks

The primary question driving this research revolves around how well native Gujarati listeners are able to distinguish two types of breathy sequences [C^{ɦ}V] and [C̤V̤] apart from all-modal sequences [CV] as well as from each other. These breathy sequences are known to have similar acoustic cues but with consistent differences in degrees and timing. In all three experimental tasks (free-sort, discrimination, and ID), however, we found that this three-way distinction was not reliably perceived by Gujarati listeners.

While the results in Sections 3–5 were separated by task type, here we look across all three tasks to reveal shared patterns in how listeners responded to stimuli representing the same phonemic group and how they responded to stimuli across groups. In Section 6.1 we discuss results of the [bar]–[bar], [b^{ɦ}ar]–[b^{ɦ}ar], and [bar]–[b^{ɦ}ar] comparisons, which suggest internally-consistent, perceptually distinct categories of the all-modal [bar] and breathy consonant [b^{ɦ}ar] categories, while in Section 6.2 we discuss the results of the [bạr]–[b̤ạr], [bar]–[b̤ạr], and [b^{ɦ}ar]–[b̤ạr] comparisons, which suggest ambiguities due to perceptual overlap of the breathy vowel [b̤ạr] category with each of the other two categories.

6.1 Robust categories and contrasts

Across tasks, results suggest that all-modal sequences [bar] and breathy consonants [b^{ɦ}ar] form categories with strong internal consistency and a robust perceptual contrast between them.

[bar] as a category: In the free-sort task, listeners were largely consistent in grouping all-modal [bar] tokens together, with the main exceptions being cases where listeners made further distinctions by talker. They were also highly accurate in categorizing [bar]–[bar] pairings as "the same" in the discrimination task and in matching the audio to an image in the ID task.

[bʱaɾ] as a category: Breathy consonant [bʱaɾ] tokens were the most consistently categorized in the free-sort task, where listeners rarely grouped them with breathy vowel [bạ̤ɾ] or all-modal [baɾ] tokens. Listeners performed significantly above chance when labeling [bʱaɾ]–[bʱaɾ] pairings as "the same" in the discrimination task, and when matching the image and audio for these tokens in the ID task.

[baɾ]–[bʱaɾ] contrast: Listeners were robustly able to keep all-modal [baɾ] and breathy consonant [bʱaɾ] distinct across tasks. They rarely grouped stimuli from these two categories together in the free-sort task, and they performed significantly above chance both in correctly identifying [baɾ]–[bʱaɾ] sequences as "different" in the discrimination task and in correctly identifying audio + image mismatches across these two categories as "different" in the ID task.

6.2 Poorly-defined categories and contrasts

Unlike with the categories in Section 6.1, listeners had much more trouble in all three tasks whenever a breathy vowel token [bạ̤ɾ] was involved. This is true both in identifying two tokens of [bạ̤ɾ] as "the same" as well as in distinguishing this category from the other two, suggesting that this category had both poor internal consistency and significant perceptual overlap with the other categories.

[bạ̤ɾ] as a category: The breathy vowel category [bạ̤ɾ] was the least consistent across all tasks. All listeners split this category across different self-sorted groupings in the free-sort task. They performed at chance in reporting two breathy vowels as "the same" in the discrimination task, and in matching the breathy vowel audio [bạ̤ɾ] with the image for /bạ̤ɾ/ 'outside' in the ID task. Taken together, this suggests that tokens of [bạ̤ɾ] might not all form a clear, distinct perceptual category for our listeners.

[baɾ]–[bạ̤ɾ] contrast: In the free-sort task, breathy vowels [bạ̤ɾ] were most frequently grouped with all-modal [baɾ]. Similarly, in the discrimination task, listeners were highly consistent in *failing* to distinguish the pairing of a breathy vowel [bạ̤ɾ] with all-modal [baɾ]. In the ID task, a bias in directionality also arose: while listeners were fairly consistent in reporting that the image for all-modal /baɾ/ 'twelve' was a *match* with the breathy vowel audio [bạ̤ɾ], they performed at chance when provided the inverse, i. e. the image for breathy vowel /bạ̤ɾ/ 'outside' paired with the all-modal audio [baɾ]. These results all point to a perceptual overlap between the all-modal [baɾ] and breathy vowel [bạ̤ɾ] categories, but not a perfect merger of the two. The bias in directionality also suggests that listeners found the breathy vowel audio [bạ̤ɾ] insufficiently breathy, since it was considered a good match with the image for /baɾ/ 'twelve',

and neither a good nor bad match with the image for /ba̤ɾ/ 'outside'; this could mean that listeners expected the audio for 'outside' to be more strongly breathy to be accepted as a match.

[bʱaɾ]–[ba̤ɾ] contrast: The free-sort task found that breathy consonant [bʱaɾ] was the most consistent group, distinguished from a common overlap between breathy vowel [ba̤ɾ] and all-modal [baɾ], and listeners were reliably able to identify the pairing of the audio for breathy vowel [ba̤ɾ] with the image for breathy consonant /bʱaɾ/ 'burden' as a mismatch, suggesting high sensitivity to differences in types of breathiness. However, the inverse of this pairing in the ID task yielded very different results, as listeners were not consistent in identifying the audio of breathy consonant [bʱaɾ] and the image of breathy vowel /ba̤ɾ/ 'outside' as a mismatch. Furthermore, listeners performed effectively at chance when judging pairings of breathy vowel [ba̤ɾ] and breathy consonant [bʱaɾ]. Like with [baɾ]–[ba̤ɾ], these results suggest some perceptual overlap between the breathy consonant [bʱaɾ] and breathy vowel [ba̤ɾ] categories, where listeners found consonantal breathiness in [bʱaɾ] to be a good representation of an underlying breathy vowel, but found vocalic breathiness [ba̤ɾ] to be subtle enough to represent all-modal /baɾ/.

7 Discussion and conclusion

In sum, results across the three tasks showed that participants do not reliably recognize the presence of breathiness in [CV̤] sequences, often indicating them to be the same as or an acceptable realization of an all-modal [CV] sequence. They do recognize the breathiness in [CʱV] sequences, but are not necessarily capable of determining whether that breathiness is associated with the vowel or consonant. Furthermore, the confusion runs in only one direction: [bʱaɾ] can be occasionally mistaken as a realization of /ba̤ɾ/, but the reverse is not true. This may shed light on the reason [ba̤ɾ] is so rarely mistaken for /bʱaɾ/: the breathiness in [ba̤ɾ] is subtle enough to pass for a fully modal /baɾ/, and not robust enough to pass for breathy consonant /bʱaɾ/. The overarching trends, then, are that (1) [CV̤] is often indistinguishable from [CV] and (2) [CʱV] can be perceived as either underlying /CʱV/ or /CV̤/.

We propose that breathiness functions like VOT in e. g. English: a continuum that is perceived categorically, with a window of ambiguity in which e. g. an alveolar stop can be perceived as either /t/ or /d/ (Eimas and Corbit 1973). The perception of breathiness can be thought of similarly but with a suite of continuous variables representing spectral tilt, spectral balance, and/or 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 [C$^{\text{ɦ}}$V] stimuli should be easily identifiable while that of [CV̤] stimuli should be more ambiguous. Consistent with the data, listeners who are sensitive to cues in degree of breathiness in [C$^{\text{ɦ}}$V] sequences, but not to cues in its timing, would be able to correctly identify a sequence as breathy, but not be able to reliably categorize it as either [C$^{\text{ɦ}}$V] or [CV̤]. Similarly, if cues to the degree of breathiness of a [CV̤] sequence are insufficient to determine with certainty that the stimulus is breathy, then that sequence would be incorrectly categorized as [CV].

A schematization of this proposal appears in Figure 7, in which the vowel after [C$^{\text{ɦ}}$] is represented with intense breathiness at first before a gradual decrease, and [V̤] is represented with more moderate, increasing breathiness. The intense breathiness associated with [C$^{\text{ɦ}}$] escapes the zone of ambiguity, while the subtle breathiness of [V̤] does not.

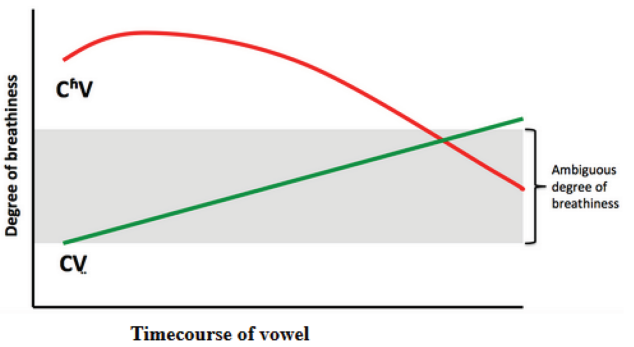

Figure 7: Schematization of degree of breathiness across timecourse of vowel.

In the scenario proposed by this explanation, listeners are sensitive to the presence or absence of intense breathiness. Significant indication of breathiness may be sufficient cause for excluding a stimulus from being modal, but insufficient cause for determining if the breathiness is associated with the consonant or vowel. The results of the present study align with this interpretation, and strongly suggest that it merits further investigation.

While ongoing work will further explore the specifics of these trends, it is evident from this study that there is a problem in differentiating [C$^{\text{ɦ}}$V] and [CV̤] sequences as well as an overlap in either the categorization or perception of [CV̤] and [CV]. These results are unexpected especially given the production evidence

demonstrating that these sounds are reliably kept distinct even in connected speech. Three observations may help shed light on the disconnect between Gujarati speakers' three-way contrast in production vs. two-way contrast in perception.

The first observation, which arguably only explains part of this mismatch, is one that has been well-established in the literature: a modal sequence [CV] may serve as an acceptable realization of a breathy vowel /CV̤/ but a breathy vowel [CV̤] is not an acceptable realization of modal /CV/. In fact, we see that in the three tasks, listeners indicated an overlap between [bə̤ɾ] and [baɾ]. Unexpectedly, however, we see in the ID task (where directionality is clearer) that listeners found the audio [bə̤ɾ] to be an acceptable realization of the image for /baɾ/ 'twelve', and not the other way around, suggesting that [bə̤ɾ] can be a realization of /baɾ/ while [baɾ] is neither reliably accepted nor rejected as a realization of /bə̤ɾ/. While the overlap is in theory expected, this reversal in directionality remains mysterious.

The second observation has to do with the low overall prevalence of the three-way contrast in the language. Recall that – due to restrictions on which vowel qualities have both modal and breathy versions across speech registers, and which vowel qualities only have both modal and breathy versions in casual registers – low vowels are arguably the primary candidate for modal vs. breathy vowel contrasts without concomitant vowel quality variation. (This restriction does not extend to breathy consonants.) Once this restriction is considered, the number of perfect minimal triplets is sharply reduced, and to our knowledge there are few minimal pairs of breathy consonant [C^{ɦ}V] vs. breathy vowel [CV̤] within a given grammatical category, reducing the weight of the already low functional load of this contrast. Thus, while these three sequences may be consistently articulated differently, listeners may not have to be attuned to such differences in perception, and can be more influenced by contextual factors. This sort of mismatch, in which speakers maintain a phonemic distinction in their production but have trouble distinguishing those sounds in perception, is widely reported in English mergers in progress (Di Paolo 1988; Herold 1990) such as *pin–pen* and *cot–caught* (Labov et al. 2005), as well as in the perception of English by Japanese listeners (Goto 1971; Sheldon and Strange 1982).

The third observation has to do with our experimental design. While our goal was to present stimuli in isolation to reduce the influence of surrounding context on syntactic, semantic, or pragmatic priming, one notable downside to this strategy is that we could not provide a "phonetic backdrop" for the word. That is to say, it is harder for listeners to normalize for the "baseline breathiness" of a given talker, e. g. how much aperiodicity in the speech should be interpreted as contrastive breathiness vs. non-phonemic characteristics

resembling breathiness in the talker's voice, or even non-speech-related background noise. This is especially problematic when distinguishing the relatively constant breathiness characteristic of breathy vowels vs. the relatively constant "modal-ness" of modal vowels. This would not, however, be an issue for breathy-aspirated stops, which inherently involve a dynamic realization of strong breathiness followed by a more modal vowel, allowing a listener to use the change in breathiness across the word as a cue. While this can help explain some of the confusions between breathy vowels and all-modal sequences, it would not independently explain why listeners were often biased to report all-modal sequences in these contexts rather than randomly guessing.

We also suggest another way to look at this: instead of focusing on the timecourse of breathiness, Gujarati listeners might be focusing on the degree of breathiness. If so, they would effectively be trying to distinguish vowels with a lack of breathiness (i. e. all-modal sequences) vs. vowels with light breathiness (i. e. breathy vowels) vs. vowels with heavy breathiness (i. e. those following breathy consonants). As mentioned above, Gujarati's (subtly) breathy vowels are acoustically most similar to the lax voice of the Yi languages, rather than to the (canonical) breathy voice of White Hmong and Valley Zapotec (Keating et al. n.d.). While breathy consonants were not included in Keating et al.'s study, we know from Esposito and Khan (2012) that they are distinguished by their significantly higher degree of breathiness compared to both breathy vowels and modal vowels. So while several languages have been documented to have either lax voice or breathy voice, Gujarati would be the only language that contrasts both of these with modal voice *and with each other*; the fact that the listeners in our study were not able to accurately distinguish the three, then, is less surprising.

These results also raise an interesting perceptual question: why are Gujarati listeners able to distinguish between modal and breathy vowels in other similar perception experiments (Bickley 1982; Esposito 2006; Kreiman et al. 2010), but not here? An answer may lie in the addition of the breathy consonants: perhaps the three-way contrast is not as salient to listeners because the addition of stimuli like [bʱaɾ] (which was excluded from the aforementioned studies), leads to higher confusability, in that the highly audible breathiness in [bʱaɾ] could be mistakenly assigned to /baɾ/. Crosslinguistically, we see that languages that contrast phonation type tend to do so on vowels or on consonants, but rarely on both. The increased perceptual difficulty in distinguishing between non-modal phonation on both consonants and vowels within a language may explain why this inventory is so typologically rare.

This preliminary account of the perception of breathy consonant vs. breathy vowels vs. modal vowels in Gujarati yields a number of avenues for future research. First, it would be important to replicate these results with a larger,

more comprehensive set of listeners, especially including more Gujarati-dominant speakers living in Gujarat. Expanding the listener population this way would allow us to investigate if there are more regions with a loss of the breathy vowel category than those previously documented. Furthermore, expanding to cover a range of ages can help us test whether the merger of breathy vowels and modal vowels is a change in progress. And, finally, now that we have examined what happens with a single minimal triplet, expanding to include more stimuli will allow us to see how lexical frequency, consonant place, and vowel quality all contribute to (and further complicate) this task.

References

Abramson, Arthur S., Theraphan L. Thongkum & Patrick W. Nye. 2004. Voice register in Suai (Kuai): An analysis of perceptual and acoustic data. *Phonetica* 61. 147–171.

Avelino, Heriberto. 2010. Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *Journal of Voice* 24. 270–280.

Berkson, Kelly H. 2013. *Phonation types in Marathi: An acoustic investigation*. Lawrence, KS: University of Kansas dissertation.

Berkson, Kelly H. 2019. Acoustic correlates of breathy sonorants in Marathi. *Journal of Phonetics* 73. 70–90.

Bickley, Corine. 1982. Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers*. 73–93. Cambridge, MA: MIT Research Lab of Electronics.

Blank, Jonah. 2002. *Mullahs on the mainframe: Islam and modernity among the Daudi Bohras*. Chicago: University of Chicago Press.

Blankenship, Barbara. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30. 163–191.

Cardona, George & Babu Suthar. 2003. Gujarati. In George Cardona & Dhanesh Jain (eds.), *The Indo-Aryan languages*, 659–697. London: Routledge.

Dave, Radhekant. 1967. A formant analysis of the clear, nasalized, and murmured vowels in Gujarati. *Indian Linguistics* 28. 1–30.

Di Paolo, Maria. 1988. Pronunciation and categorization in sound change. In Kathleen Ferarrara, Becky Brown, Keith Walters & John Baugh (eds.), *Linguistic change and contact: NWAV-XVI*, 84–92. Austin: Department of Linguistics, University of Texas.

DiCanio, Christian T. 2010. Itunyoso Trique. *Journal of the International Phonetic Association* 40 (2). 227–238.

Dixit, R. Prakash. 1989. Glottal gestures in Hindi plosives. *Journal of Phonetics* 17. 213–237.

Dutta, Indranil. 2007. *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*. Urbana & Champaign, IL: University of Illinois Urbana-Champaign dissertation.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.) 2019. Gujarati. *Ethnologue: Languages of the world*, 22nd edn. Dallas: SIL International. Online version. http://www.ethnologue.com.

Eimas, Peter & John Corbit. 1973. Selective adaptation of linguistic feature detectors. *Cognitive Psychology* 4(1). 99–109.
Esposito, Christina M. 2006. *The effects of linguistic experience on the perception of phonation*. Los Angeles: University of California Los Angeles dissertation.
Esposito, Christina M. 2010a. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38(2). 303–316.
Esposito, Christina M. 2010b. Variation in contrastive phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association* 40(2). 181–198.
Esposito, Christina M. & Sameer ud Dowla Khan. 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.
Esposito, Christina M. & Sameer ud Dowla Khan. To appear (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass*.
Fischer-Jørgensen, Eli. 1967. Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics* 28. 71–139.
Garellek, Marc. 2019a. The phonetics of voice. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics*. Philadelphia: Routledge.
Garellek, Marc. 2019b. Acoustic discriminability of the complex phonation system in !Xóõ. *Phonetica*. doi: 10.1159/000494301.
Garellek, Marc & Patricia A. Keating. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41(2). 185–205.
Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29. 383–406.
Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R.". *Neuropsychologia* 9(3). 317–323.
Granqvist, Svante. 2003. *The visual sort and rate method for perceptual evaluation in listening tests*. *Logopedics Phoniatrics Vocology* 28(3). 109–116. doi: 10.1080/14015430310015255.
Herold, Ruth. 1990. *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania*. Philadelphia: University of Pennsylvania dissertation.
Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America* 121. 2283–2295.
Jones, Ted E. & Lyle M. Knudson. 1977. Guelavia Zapotec phonemes. In William Merrifield (ed.), *Studies in OtoManguean phonology*, 163–180. Arlington, Dallas: SIL/University of Texas.
Keating, Patricia A., Christina M. Esposito, Marc Garellek, Khan Sameer ud Dowla & Jianjing Kuang. n.d. A cross-language acoustic space for phonation distinctions. Unpublished manuscript.
Key, Michael P. 2012. *Phonological and phonetic biases in speech perception*. Amherst, MA: University of Massachusetts, Amherst dissertation.
Khan, Sameer ud Dowla. 2010. Bengali (Bangladeshi standard). *Journal of the International Phonetic Association* 40(2). 221–225.
Khan, Sameer ud Dowla. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40. 780–795.
Kirk, Paul L., Jenny Ladefoged & Peter Ladefoged. 1993. Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec. In Anthony Mattina & Timothy

Montler (eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, 435–450. Missoula: University of Montana Press.
Kreiman, Jody, R. Bruce & Sameer ud Dowla Khan. 2010. Effects of native language on perception of voice quality. *Journal of Phonetics* 384. 588–593.
Labov, William, Sharon Ash & Charles Boberg. 2005. *The atlas of North American English: Phonetics, phonology, and sound change*. Berlin: De Gruyter.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.
Laver, John. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
Miller, Amanda L. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35. 56–84.
Mistry, P. J. 1997. Gujarati phonology. In Alan S. Kaye (ed.), *Phonologies of Asia and Africa*, 653–673. Winona Lake: Eisenbrauns.
Modi, Bharati. 1986. Rethinking of "murmur in Gujarati". *Indian Linguistics* 47. 39–55.
Munro, Pamela, Felipe H. Lopez, Olivia Méndez, Rodrigo Garcia & Michael R. Galant. 1999. *Di'syonaary X:tee'n Diizh Sah Sann Lu'uc (San Lucas Quiavinî Dictionary/Diccionario Zapoteco de San Lucas Quiavinî)*. Los Angeles: Chicano Studies Research Center Publications.
Nair, Usha. 1979. *Gujarati phonetic reader*. Mysore: Central Institute of Indian Languages.
Nara, Kiranpreet. 2017. Acoustic and electroglottographic study of breathy and modal vowels as produced by heritage and native Gujarati speakers. *Proceedings of Interspeech 2017*, 1054–1058.
Nî Chasaide, Ailbhe & Christer Gobl. 1997. Voice source variation. In William J. Hardcastle & John Laver (eds.), *The handbook of phonetic sciences*, 427–461. Oxford: Blackwell.
Ohala, Manjari. 1983. *Aspects of Hindi phonology*. Delhi: Motilal Banarsidass.
Pandit, Prabodh Bechardas. 1957. Nasalization, aspiration and murmur in Gujarati. *Indian Linguistics* 17. 165–172.
Rami, Manish K., Joseph Kalinowski, Andrew Stuart & Michael P. Rastatter. 1999. Voice onset times and burst frequencies of four velar stop consonants in Gujarati. *Journal of the Acoustical Society of America* 106(6). 3736–3738.
Sheldon, Amy & Winifred Strange. 1982. The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3(3). 243–261.
Shue, Yen-Liang, Patricia A. Keating, Chad Vicenik & Yu Kristine. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1846–1849.
Traill, Anthony. 1985. *Phonetic and phonological studies of !Xóõ Bushman*. Hamburg: Helmut Buske.
Watkins, Justin. 1999. CQ of laryngeal gestures and settings in Wa. *Proceedings of the 14th International Congress of Phonetics Sciences*, 1017–1020.
Watkins, Justin. 2002. *The phonetics of Wa: Experimental phonetics, phonology, orthography and sociolinguistics*. Canberra: Pacific Linguistics.
Yadav, Ramawatar. 1984. Voicing and aspiration in Maithili: A fiberoptic and acoustic study. *Indian Linguistics* 45. 1–30.