DISTINGUISHING BREATHY CONSONANTS AND VOWELS IN GUJARATI

christina m. esposito, macalester college
sameer ud dowla khan, reed college
kelly h. berkson, univ. indiana bloomington
max nelson, univ. massachusetts amherst

icphs-19, melbourne, 9 august 2019
DISTINGUISHING BREATHY CONSONANTS AND VOWELS IN GUJARATI

christina m. esposito, macalester college
sameer ud dowla khan, reed college
kelly h. berkson, univ. indiana bloomington
max nelson, univ. massachusetts amherst

icphs-19, melbourne, 9 august 2019
Big question

Several lgs have contrastive **breathiness** on:

- **Consonants** ($C^\text{ți}$)

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>V vs. V</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>English, Punjabi, Std. Khmer</td>
<td></td>
</tr>
<tr>
<td>C vs. $C^\text{ți}$</td>
<td><strong>Bengali</strong> (Khan 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[gaʃ] ‘you sing’</td>
<td>[g^țiaʃ] ‘grass’</td>
</tr>
</tbody>
</table>
Big question

- Several lgs have contrastive breathiness on:
  - **Consonants** ($C^{\text{fi}}$), or
  - **Vowels** ($\text{V}$)

- **W. Khmer** (Wayland & Jongman 2003)
  - [pɔː:k] ‘by chance’
  - [pɔː:k] ‘bumped’

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>V vs. $\text{V}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$ vs. $C^{\text{fi}}$</td>
<td>Bengali Javanese Tsonga</td>
<td></td>
</tr>
</tbody>
</table>
Big question

- But very **few have both**: why?
- Is it that $C^h$ has to express its breathiness on adjacent $V$, generating confusion with $CV$?

**Gujarati** (Esposito & Khan 2012)

<table>
<thead>
<tr>
<th>V</th>
<th>V vs. V</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$ vs. $C^h$</td>
<td>Bengali Javanese Tsonga</td>
</tr>
</tbody>
</table>
Gujarati breathiness

- Well-known minimal triplet of Gujarati:
  - All modal \([\bar{b}aɾ]\) ‘twelve’
  - Breathy \(\check{C}^{\tilde{a}}\) \([b^{\tilde{a}}aɾ]\) ‘burden’
  - Breathy \(V\) \([\bar{b}aɾ]\) ‘outside’
Narrow question

- In their production, Gujarati speakers distinguish \( C^hV \) vs. \( CV \) vs. \( CV \) (Esposito & Khan 2012)
  - True even for heritage speakers (Nara 2017)
- They are also highly sensitive to tiny variations in \( H1-H2 \) in perception (Kreiman, Gerratt, & Khan 2010)
  - True even for heritage speakers
  - But not tested for 3-way comparison
- So can speakers use this sensitivity to distinguish \( C^hV \) vs. \( CV \) vs. \( CV \) in perception?
Predictions

- $C^fV$ & $CV$ use the **same cues** of breathiness (Khan 2012; Dave 1967; Fischer-Jørgensen 1967)
  - **Steeper spectral slope**, measured as higher $H1-H2$, $H1-A1$, $H1-A2$, $H1-A3$, $H2-H4$
  - **Noisier**, measured as lower CPP, HNR
- Suggests $C^fV$ and $CV$ are confusable
Predictions

- But $C^hV$ & $CV$ differ in cue realization (Esposito & Khan 2012)
  - $C^hV$ has a shorter, more extreme breathiness
  - $CV$ has longer, more subtle breathiness
- Suggests $C^hV$ and $CV$ can be distinguished

![Graph showing differences in breathiness between $C^hV$, $CV$, and $CV^\sim$ across time points]
## Predictions

- $C^\text{fi}$ and $V$ are not equally “stable” (Cardona & Suthar 2003; Mistry 1997; Modi 1987; Nair 1979; Dave 1967; Pandit 1957)

- $V$ may be especially susceptible to merger w/ $V$

<table>
<thead>
<tr>
<th></th>
<th>Breathy $C^\text{fi}$</th>
<th>Breathy $V$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regionally</strong></td>
<td>Stable</td>
<td>$V \rightarrow V$ in some varieties</td>
</tr>
<tr>
<td><strong>Orthographically</strong></td>
<td>Dedicated graphemes</td>
<td>Represented with $\varepsilon &lt;f_i&gt;$ or not at all</td>
</tr>
<tr>
<td><strong>In careful speech</strong></td>
<td>Stable</td>
<td>$V \rightarrow Vf_iV$</td>
</tr>
<tr>
<td><strong>In fast speech</strong></td>
<td>$C^\text{fi}V \rightarrow CV$</td>
<td>$V \rightarrow V$</td>
</tr>
<tr>
<td></td>
<td>$VC^\text{fi} \rightarrow VC$</td>
<td></td>
</tr>
</tbody>
</table>

$u^C^\text{fi}$ and $V^\text{fi}$ are not equally “stable” (Cardona & Suthar 2003; Mistry 1997; Modi 1987; Nair 1979; Dave 1967; Pandit 1957)

$V^\text{fi}$ may be especially suscep+ble to merger w/ $V$
Overview of methods

- 3 perception tasks
  - Free-sort
  - AX discrimination
  - Picture-matching

- All tasks used the same minimal triplet stimuli, taken from production study (Khan 2012)
  - Naturally-produced
  - Excised from connected speech
  - Utterance-initial

More open-ended

More lexically-determined
### Participants

<table>
<thead>
<tr>
<th>Talker #</th>
<th>Sex</th>
<th>Age</th>
<th>Home city</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>F</td>
<td>22</td>
<td>Mumbai, Maharashtra, India</td>
</tr>
<tr>
<td>T2</td>
<td>F</td>
<td>20s</td>
<td>Mumbai, Maharashtra, India</td>
</tr>
<tr>
<td>T3</td>
<td>F</td>
<td>23</td>
<td>Mumbai, Maharashtra, India</td>
</tr>
<tr>
<td>T4</td>
<td>F</td>
<td>30</td>
<td>Mumbai, Maharashtra, India</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Listener #</th>
<th>Sex</th>
<th>Age</th>
<th>Birthplace</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>M</td>
<td>23</td>
<td>Ahmedabad, Gujarat, India</td>
</tr>
<tr>
<td>L2</td>
<td>M</td>
<td>30</td>
<td>Vadodara, Gujarat, India</td>
</tr>
<tr>
<td>L3</td>
<td>M</td>
<td>23</td>
<td>Mumbai, Maharashtra, India</td>
</tr>
<tr>
<td>L4</td>
<td>M</td>
<td>26</td>
<td>Ahmedabad, Gujarat, India</td>
</tr>
<tr>
<td>L5</td>
<td>F</td>
<td>52</td>
<td>Bardoli, Gujarat, India</td>
</tr>
<tr>
<td>L6 (heritage)</td>
<td>F</td>
<td>19</td>
<td>Fostoria, Ohio, USA</td>
</tr>
</tbody>
</table>
Task 1: free sort

- 24 audio stimuli presented as visual icons
  - Could click & listen as many times as desired
Task 1: free sort

- 24 audio stimuli presented as visual icons
  - Could click & listen as many times as desired
  - Listeners asked to **drag and sort** into as many groups as desired (2+)
  - Not given any further instruction
Task 1: free sort

- Listeners 3 & 4 made:
  - a [b̥ar] group

Listener 4

Listener 3
Task 1: free sort

- Listeners 3 & 4 made:
  - a [b̥ar] group
  - a [bar] + [b̥ar] group
  - suggests perceptual overlap

Listener 4

Listener 3
Task 1: free sort

- Listener 5 made three groups
- Roughly the three lexical items, with lots of mixing

**Listener 5**
Task 1: free sort

- Listeners 1 & 2 went further
- Grouped by word and talker
- Suggests highly sophisticated auditory skill

**Listener 1**

**Listener 2**
Task 1: free sort

- Listener 6 made groups we could not interpret
Task 2: AX discrimination

- Classic **AX discrimination** task
  - Listeners heard all 54 pairs of stimuli
  - No pair had the same talker for both words
  - ISI=300ms, response time <1000ms, no rep.
  - Clicked સામાનુક ‘same’ or મોટુક ‘different’

- Results displayed as
  - Bars for % correct
  - Chi-square tests for significance
Task 2: AX discrimination

In “same” trials:

- [bar]-[bar] was clearly “same”
Task 2: AX discrimination

- In “same” trials:
  - [bar]-[bar] was clearly “same”
  - [b̃ar]-[b̃ar] was clearly “same”
Task 2: AX discrimination

- In “same” trials:
  - [bar]-[bar] was clearly “same”
  - [bʰar]-[bʰar] was clearly “same”
  - [bar]-[bar] responses were not sig. different from chance
- Suggests [bar] is hard to identify as a single category
Task 2: AX discrimination

- In “different” trials:
  - [\text{bar}] - [b^6\text{ar}] was “different”
Task 2: AX discrimination

- In “different” trials:
  - [bar]-[bʰar] was “different”
  - [b̥ar]-[bʰar] was not sig. different from chance
Task 2: AX discrimination

- In “different” trials:
  - [b̃ar]-[b̃ʱar] was “different”
  - [b̃ar]-[b̃ʱar] was not sig. different from chance
  - [b̃ar]-[bar] was “same”!

- Suggests overlap of [b̃ar] with [bʱar], and merger with [bar]
Task 3: picture-matching

- **ID task with images** representing words
  - Audio followed by image
  - Clicked 'same' or 'different'
  - This was the task most strongly determined by lexical categories

- All modal
  - [bar] ‘twelve’

- Breathy Cʰ
  - [bʰar] ‘burden’

- Breathy V̆
  - [b̆ar] ‘outside’
Task 3: picture-matching

- Results in the form of a confusion matrix
- If listeners are great at this task:
  - Expect $* >$ chance in grey boxes
  - Expect $* <$ chance elsewhere

<table>
<thead>
<tr>
<th>Image</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>/baɾ/  ‘twelve’</td>
<td>[baɾ]</td>
</tr>
<tr>
<td>/bʰaɾ/ ‘burden’</td>
<td></td>
</tr>
<tr>
<td>/b̥aɾ/ ‘outside’</td>
<td></td>
</tr>
</tbody>
</table>
Task 3: picture-matching

- [bar] and [bʱar] were robust, distinct groups
  - Audio+image matches consistently accepted
  - Mismatches consistently rejected

<table>
<thead>
<tr>
<th>Image</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>/bar/ 'twelve'</td>
<td>97.5*</td>
</tr>
<tr>
<td>/bʱar/ 'burden'</td>
<td>17.5*</td>
</tr>
<tr>
<td>/b̄r/ 'outside'</td>
<td></td>
</tr>
</tbody>
</table>
**Task 3: picture-matching**

- [baɾ] however, was problematic throughout
- Listeners did not consistently accept or reject any audio paired with /baɾ/ image

<table>
<thead>
<tr>
<th>Image</th>
<th>Audio</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/baɾ/ ‘twelve’</td>
<td>[baɾ]</td>
<td>97.5*</td>
<td>5.0*</td>
</tr>
<tr>
<td>/b̥aɾ/ ‘burden’</td>
<td>[b̥aɾ]</td>
<td>17.5*</td>
<td>70.0*</td>
</tr>
<tr>
<td>/b̥aɾ/ ‘outside’</td>
<td>[baɾ]</td>
<td>65.0</td>
<td>42.5</td>
</tr>
</tbody>
</table>
Task 3: picture-matching

- Listeners rejected [bær] audio for /b̥ar/ image.
- They accepted [bær] audio for /bar/ image (!)
- And they were inconsistent about whether [bær] audio matched the /b̥ar/ image itself.

<table>
<thead>
<tr>
<th>Image</th>
<th>Audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>/bar/ ‘twelve’</td>
<td>[bar] 97.5*</td>
</tr>
<tr>
<td>/b̥ar/ ‘burden’</td>
<td>17.5*</td>
</tr>
<tr>
<td>/b̥ar/ ‘outside’</td>
<td>65.0</td>
</tr>
</tbody>
</table>
Summary of results

- **Two robust categories**
  - All modal [b̥aɾ]:
    - fairly consistently grouped in free-sort
    - 87.5% “same” in audio pairs
    - 97.5% “same” when paired with own image
  - Breathy consonant [bʱaɾ]:
    - very consistently grouped in free-sort
    - 83.3% “same” in audio pairs
    - 70.0% “same” when paired with own image
Summary of results

- **One robust contrast**

- [bær] vs. [bʰær]:
  - rarely grouped in free-sort
  - 78.5% “different” in audio pairs
  - 82.5% “different” w/ opposite image
Summary of results

- One poorly-defined category
- Breathy vowel [bəɾ]:
  - most inconsistently sorted
  - neither “same” nor “different” in audio pairs
  - neither “same” nor “different” w/ own image
Summary of results

- **Two poorly-defined contrasts**
  - [ɓ̥ɑɾ] vs. [b̥ɑɾ]:
    - rarely collapsed in free-sort
    - BUT inconsistent responses for audio pairs
  - [b̥ɑɾ] vs. [bar]:
    - most likely to collapse in free-sort
    - 68.1% “same” in audio pairs
- Directionality: [ɓ̥ɑɾ] audio rejected for /b̥ɑɾ/ image, and accepted for /bar/ image… but inconsistent responses for inverse pairings
Conclusions

- Gujarati spkrs produce a 3-way breathiness distinction, but they **do not reliably perceive it**
- **Subtle, constant breathiness** of $CV^\gamma$ is **missed**
- **Robust, brief breathiness** of $C^nV$ is perceived…
- …but is often **mis-associated to $CV^\gamma$**

- **New question**: how are they producing this contrast if they aren’t clearly perceiving it?
- More data is needed
- Before expanding study further, we’d love to hear your feedback!
Acknowledgments

To my co-authors, our talkers, our listeners, and to everyone in the audience here at ICPhS-19:

[abha]  
आभार  
‘thanks’
References


Gujarati breathiness

4-way contrast in **consonantal glottal state**

<table>
<thead>
<tr>
<th>Orthography</th>
<th>IPA</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>-voi -asp</td>
<td>kal</td>
<td>‘yesterday/tomorrow’</td>
</tr>
<tr>
<td>-voi +asp</td>
<td>kʰal</td>
<td>‘drain’</td>
</tr>
<tr>
<td>+voi -asp</td>
<td>gal</td>
<td>‘curse word’, ‘filter!’</td>
</tr>
<tr>
<td>+voi +asp</td>
<td>gʰal</td>
<td>‘penetrate!’</td>
</tr>
</tbody>
</table>

2-way contrast in **vocalic glottal state**

<table>
<thead>
<tr>
<th>Orthography</th>
<th>IPA</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>modal</td>
<td>kan</td>
<td>‘ear’</td>
</tr>
<tr>
<td>breathy</td>
<td>kᵃn(o)</td>
<td>‘Krishna’</td>
</tr>
</tbody>
</table>
Predictions

- Distinguishing CV, CV^̅, CV^ using H1–H2

![Graph showing predictions for CV, CV^ and CV^̅ using H1-H2. The graph includes timepoints 1 to 5 and three curves representing different conditions. The conditions are labeled as CV[bar], CV[bhar], and CV^[bar].]
Why just these target words?

- Task might be difficult enough, so a minimal triplet is desirable
- Roughly equal lexical frequency
- /b/ → [ɓ] optionally before modal but not breathy Vs, which can aid listeners (Vyas 1978)
- Breathiness is most perceptible in low /a/ (Fischer-Jørgensen 1967)
- Mid /ɛ ɔ/ tend to lower to /ɛ ɔ/ when breathy (Dave 1967)