ARTICLE IN PRESS

Journal of Phonetics ▪ (▪▪▪▪) ▪▪▪–▪▪▪

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/phonetics

# The effects of linguistic experience on the perception of phonation

Christina M. Esposito *

*Macalester College, St. Paul, MN 55105-1801, USA*

ARTICLE INFO

*Article history:*
Received 7 April 2009
Received in revised form
7 February 2010
Accepted 10 February 2010

ABSTRACT

This study investigates the role linguistic experience has on the perception of phonation and acoustic properties that correlate with this perception. Listeners from Gujarati (contrasts breathy versus modal vowels), Spanish (no breathiness) and English (allophonic breathiness) participated in: (1) a similarity-rating task, indicating the similarity of modal and/or breathy Mazatec vowels and (2) a free-sort task, sorting breathy and modal stimuli from many languages.

Results showed that Gujaratis did better at distinguishing phonation in other languages/dialects and were more consistent. English listeners did no better than Spanish listeners, despite the allophonic breathiness in English. In terms of acoustic dimensions, results showed that Gujaratis relied on H1 – H2 (amplitude of the first harmonic minus amplitude of the second harmonic), English listeners relied weakly on H1 – H2 and cepstral peak prominence and Spanish listeners relied on H1 – A1 (amplitude of first formant peak) and H1 – H2. While it is not clear why Spanish listeners used H1 – A1, we can speculate as to why all three groups of listeners used H1 – H2. Cross-linguistically, H1 – H2, which is correlated with the open quotient (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995), is the most successful measure of phonation. Perhaps the reason is a perceptual one; open quotient differences might be more salient to listeners.



## 1. Introduction

Most of the phonetic research on linguistically relevant phonation (that is, phonation used either phonemically or allophonically in a language, as opposed to pathologically disordered phonations) involves experimental descriptions of the production of phonation in particular languages. Examples include: Fischer-Jørgensen (1967) on Gujarati, Huffman (1987) on Hmong, Thongkum (1987) on Nyah Kur and Kui, Kirk, Ladefoged, and Ladefoged (1993) on Jalapa Mazatec and, more recently, Andruski and Ratliff (2000) on Green Mong, Wayland and Jongman (2003) on Khmer, Abramson, Thongkum, and Nye (2004) on Suai, and Esposito (2003, 2004) on Zapotec, to name a few. Other studies have expanded this research to include specific properties of phonation such as the localization of non-modal phonation (Silverman, 1997), the duration of non-modal phonation (Blankenship, 1997), the relationship between tongue root position and phonation (Guion, Post, & Payne, 2004), and the interaction between intonation and phonation (Epstein, 2002; Esposito, 2003). These and other similar studies have greatly contributed to our understanding of the production of phonation. However, there is still little known about the perception of phonation.

Most of what we do know about the perception of phonation comes from studies where English listeners judge pathologically disordered voice qualities (e.g. Kreiman & Gerratt, 1996; Kreiman, Gerratt, Antonanzas-Barroso, & Berke, 1993; Kreiman, Gerratt, & Precoda, 1990), a task they often do quite poorly on, with little inter-subject consistency. Additional studies have examined the perception of phonation which is neither linguistically relevant nor pathologically disordered. For example, Klatt and Klatt (1990), using synthesized stimuli, found that aspiration noise, rather than increased amplitude of the first harmonic (H1), was perceptually the most important cue to voice quality variation for English listeners. In addition, Hillenbrand, Cleveland, and Erickson (1994) using modal and breathy phonation (as produced by trained English speakers), found that cepstral peak prominence (CPP) was the best predictor of perceived breathiness for English listeners.

While the aforementioned studies have made much progress toward an understanding of the perception of phonation by English listeners, less is known about the perception of phonation by listeners with phonation contrasts. There have been three studies on the topic: Fischer-Jørgensen (1967), Bickley (1982), and Gerfen and Baker (2005). Fischer-Jørgensen (1967) and Bickley (1982) both studied the perception of phonation by Gujaratis (a language that contrasts breathy and modal vowels) and found that the amplitude of the first harmonic (H1) played an important role in perception. More specifically, both Fischer-Jørgensen (1967) and Bickley (1982) found that Gujarati listeners successfully identified breathy vowels, especially in cases where the spectrum was dominated by H1. In particular, in Bickley's study, Gujarati listeners found vowels with H1 – H2 (where H2 is the amplitude of the second harmonic) values of approximately 12.5 dB to be 'very breathy', vowels with an average H1 – H2 value

* Tel.: +1 651 696 6678; fax: +1 651 333 4754.
*E-mail address:* esposito@macalester.edu

0095-4470/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2010.02.002

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002

of 8.3 dB to be 'breathy', vowels with an average H1 – H2 value of 6.7 dB to be 'slightly breathy', and vowels with an average H1 – H2 value of 0 dB to be 'not breathy'.

An additional study on both the production and perception of phonation, Gerfen and Baker (2005), showed that, in production, Mixtec laryngealized vowels were distinguished from modal ones by four acoustic phonetic properties: a drop in fundamental frequency (f0), a drop in amplitude, vowel duration, and H1 – H2. They tested the role of f0 and amplitude separately (but not of vowel duration or of H1 – H2) in the perception of modal versus laryngealized phonation. Results showed that either a drop in f0 or a drop in amplitude could be used by listeners to identify phonation contrasts in the absence of spectral cues. Taken together these three studies represent our knowledge of the perception of phonation by listeners of languages that contrast phonation.

The perception of phonation by listeners of languages that contrast phonation is especially interesting when considering the fact that in a language with contrastive phonation there can be a wide variety of productions. For example, in !Xóõ breathy vowels are distinguished from modal ones by open quotient differences (as indicated by H1 – H2), spectral tilt, and/or noise, depending on the speaker (Ladefoged & Antoñanzas-Barroso, 1985). Presumably, listeners must be able to associate more than one acoustic property with a particular phonation category and weigh these acoustic properties accordingly.

The goal of the present study is to expand our current understanding of the perception of linguistically relevant phonation by examining the role linguistic experience plays in the perception of phonation and the acoustic substrates of this perception. More specifically, three groups of listeners, English (which has allophonically breathy vowels), Gujarati (which contrasts breathy and modal vowels, phonemically) and Spanish (which is reported to only have modal vowels), were asked to participate in two tasks: (1) a similarity-rating task using stimuli drawn from Mazatec, a language that contrasts breathy, modal, and creaky phonation and (2) a free-sort task using stimuli drawn from a wide variety of languages/dialects that differ in their production of phonation. Taken together, the experiments in this study act like a categorical perception task. The similarity-rating task, while not a straightforward categorical perception task, can be interpreted as one. If listeners' responses fall into clusters in the similarity-rating task, these clusters could be interpreted as categories. The nature of these categories can be further explored in the free-sort task which is like an identification task, differing only in the nature of the overt response. (Speakers were not told what categories were available to them, only that there were two categories, into which they had to sort the stimuli.)

It is hypothesized that, due to their phonemic phonation contrast, Gujarati listeners will be better at distinguishing modal from breathy phonation and that they will perform more consistently than either English or Spanish listeners. In addition, it is hypothesized that the strongest acoustic correlate(s) of phonation in a given language (which will be discussed for each language in Section 2) will be the most salient acoustic property for listeners of that language.

## 2. Phonation in English, Gujarati, and Mexican Spanish

This section will review the allophonic or phonemic phonation types found in the languages spoken by the listener population, English, Gujarati and Mexican Spanish.

### 2.1. English

Phonemically, English only has modally phonated vowels. However, there are cases of allophonic creaky or breathy voice: (1) creakiness at the beginning of vowel-initial words associated with an allophonic glottal stop (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996), (2) creakiness associated with the ends of sentences and paragraphs (Kreiman, 1982), (3) non-modal phonation at phrase boundaries (Epstein, 2002; Pierrehumbert & Talkin, 1992) and focal pitch-accented words (Epstein, 2002), and (4) breathiness on vowels after /h/ (Epstein, 1999; Ladefoged, 1983; Lofqvist & McGowan, 1992). Epstein (1999) showed that the allophonic breathiness in English is reflected in the open quotient, which is associated with the acoustic measure H1 – H2 (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995).

### 2.2. Gujarati

Gujarati is unusual among the Indic languages in having a breathy versus modal vowel distinction, in addition to a breathy versus modal voiced consonant distinction. Fischer-Jørgensen (1967) found that one of the most prominent features distinguishing Gujarati breathy from modal vowels was a dominant H1; on average, H1 was 4.37 dB higher for breathy vowels than for modal ones. Gujarati was chosen because it is one of the very few languages that has a phonation contrast on vowels without phonetic tone.

### 2.3. Mexican Spanish

Like English, Mexican Spanish does not have contrastive phonation, and phonemically has modally phonated vowels. However, Mexican Spanish differs from English in that no study of Spanish has described discourse or allophonic phonation variation of the sort described in English. To verify this, 10 native speakers of Mexican Spanish were recorded producing 10 sentences of different types (e.g. declarative, interrogative, imperative, etc.). Spectrograms were made of each of these sentences and were visually inspected for any signs of breathiness. If there were any signs of breathiness (i.e. a decrease in higher frequency energy, visible noise, etc.) spectral measures were made to determine the phonation of the segment in question. For all of the sentences recorded, only modal phonation was present. In addition, recordings made by Andrade (2003) for a study on the intonation of Mexican Spanish were examined in the same manner. Breathy phonation was not observed. However, in the absence of large scale studies on phonation, it is still possible that Mexican Spanish does contain allophonic non-modal phonation. Furthermore, there are some dialects with potential breathiness. For example, in the southeastern dialects of Mexican Spanish, /x/ is replaced with [h], and in the area around Veracruz, /s/ is debuccalized (Canfield, 1981).

## 3. Experiment 1: similarity-rating task

In this section, methodology and results of the similarity-rating task are presented. (This research was originally part of a larger project that contained three randomly ordered similarity-rating tasks. The results across all three rating tasks were very similar. In the interest of space, only one of the three will be presented here. For the others, see Esposito, 2006.)

### 3.1. Methods

#### 3.1.1. Listeners

*3.1.1.1. English.* A total of 18 native speakers of American English participated in this experiment. Of the 18 listeners, eight were male; 10 were female. None of the listeners had any experience with languages with phonemically non-modally phonated vowels, and none of the listeners were familiar with the purpose of the study. All listeners were undergraduate students at the University of California, Los Angeles, and all received extra credit in an introductory linguistics course for their participation.

*3.1.1.2. Gujarati.* Twelve native speakers of Gujarati participated in this experiment; none were familiar with the purpose of the study. Of the 12 listeners, six were male; six were female. All of the listeners were born in Gujarat, India, but now live in the US and speak Indian English and Hindi fluently in addition to Gujarati. Each subject received $10 for his/her participation.

*3.1.1.3. Spanish*

Eighteen native speakers of Mexican Spanish participated in this experiment. Of the 18 listeners, seven were male; 11 were female. All of the subjects were born in Mexico, but now live in the US, and all speak English as a second language to varying extents. None of the listeners had any experience with languages with non-modally phonated vowels, except for English. Native speakers of dialects with potential breathiness were avoided. None of the listeners were familiar with the purpose of the study. All of the listeners were undergraduate students at the University of California, Los Angeles, and received either $10 or extra credit in an introductory linguistics course for their participation.

*3.1.2. Stimuli*

Naturally produced stimuli were drawn from Jalapa Mazatec (hereafter Mazatec), an Otomanguean language spoke in San Felipe Jalapa Díaz, Oaxaca, Mexico. Mazatec was chosen for a number of reasons: it has a phonation contrast on vowels (breathy, modal, and creaky), its tones and phonations are fully cross-classified, it has been the subject of studies on phonation (e.g. Blankenship, 1997; Silveram et al., 1995; Kirk et al., 1993), and extensive recordings were available from the UCLA Phonetics Archive. Mazatec breathy and modal vowels are best distinguished by the measures H1 – H2, H1 – A2, and CPP (Blankenship, 1997). (Because creaky phonation is not relevant to this study, it will be ignored here.) The recordings used in this study were made by Paul Kirk and Peter Ladefoged in 1993. The recordings include six male and six female speakers producing words, including some minimal sets, in isolation.

*3.1.2.1. Manipulation of the stimuli.* The stimuli were composed of eight breathy and eight modal vowels excised from real words consisting of a coronal stop followed by [a]. (Coronals were chosen because there were more tokens available for this place of articulation.) Two breathy and two modal vowels were chosen from each language, producing a total of 40 stimuli. The vowels were cut right after the stop burst, leaving the CV formant transition intact. This created a language-neutral stop-like sound at the beginning of the vowel, while crucially preserving any important phonation cues. In addition, cutting the vowel right after the stop burst avoided a problematic rise in the amplitude at the onset of the vowel, which could be perceived as a glottal stop.

Each vowel was normalized to an average duration of 250 ms by copying and pasting (or cutting, when natural vowel duration exceeded 250 ms) individual pulses. (The original duration of the vowels ranged from 190 to 310 ms.) The original proportion of breathy to modal phonation was always preserved. F0 was normalized to a slightly falling f0 of 160–140 Hz using the PSOLA (pitch-synchronous overlap and add) function of Praat software. PSOLA changes the f0 of a signal without changing other properties of the voice. (For a full explanation of how this works, see Upperman, 2004.) Even though PSOLA, in theory, does not change the spectrum, to err on the side of caution, the f0 of the stimuli was re-synthesized within a range of $\pm 40$ Hz from the original value. (For example, if the original f0 was 100 Hz, it would be re-synthesized to a value in the range of 60–140 Hz.) This is because pilot data showed that re-synthesizing within this range created a $<1$ dB change in the harmonics (Esposito, 2005). (Expert listeners from the UCLA Phonetics Lab could not hear the $<1$ dB difference between the natural tokens and the re-synthesized ones, Esposito, 2005.)

*3.1.2.2. Phonetic measurements of stimuli.* Seven acoustic measures were made on each stimulus to determine which dimension(s) might correlate best with listeners' perception. (Only acoustic measures were made because the stimuli were only available as audio signals.) Of these seven measures, one measure, cepstral peak prominence (CPP), was a measure of periodicity. Cepstral peak prominence was taken automatically using software based on Hillenbrand et al. (1994) and created by the Bureau of Glottal Affairs, Division of Head/Neck Surgery, UCLA School of Medicine. The other six were spectral measures made by hand from a Fast Fourier Transform. These included the difference between the amplitudes of the:

(1) first and second harmonics (H1 – H2),
(2) first harmonic and the principal harmonic of the first formant (H1 – A1),
(3) first harmonic and the principal harmonic of the second formant (H1 – A2),
(4) first harmonic and the principal harmonic of the third formant (H1 – A3),
(5) the principal harmonic of the second and third formants (A2 – A3)
and
(6) the average of H1 + H2 minus A1 (((H1 + H2)/2) – A1).

Theses seven measures were selected because they are among the most common measures of phonation used in previous studies on phonation (H1 – H2: e.g. Bickley, 1982; Blankenship, 1997; Fischer-Jørgensen, 1967; Huffman, 1987; H1 – A3: Blankenship, 1997; Stevens & Hanson, 1995; H1 – A1 or H1 – A2: Ladefoged, 1983; the average of H1 + H2 compared to A1: Stevens, 1988 and A2 – A3: Klatt & Klatt, 1990). For a review of these measures see, Gordon and Ladefoged (2001).

*3.1.2.2.1. Results of stimuli measurements.* Fig. 1 is a set of graphs of CPP, H1 – H2, H1 – A1, H1 – A2, H1 – A3, A2 – A3 the average of H1 – H2 compared to A1, for the 16 stimuli. In the graphs, the *y*-axis is in dB.

Results showed that CPP, H1 – H2, H1 – A1, and H1 – A2 are good measures of phonation for the Mazatec stimuli in that they successfully distinguished the modal and breathy phonation. A measure was considered successful if there was a significant difference between the phonations in any direction. Significance was based on two-tailed *t-tests*.

Discriminant analysis, a procedure that determines which variables discriminate between two or more groups, was used to determine if these seven measures were sufficient to distinguish breathy from modal Mazatec stimuli. The results of the discriminant analysis showed that these seven acoustic measures accounted for 87% of the variance in the data, with H1 – A2 accounting for most of the variance (53%), followed by H1 – A1 (20%) and then H1 – H2 (14%).

*3.1.3. Procedure*

Each listener participated in the similarity-rating task, which was randomly ordered with respect to the free-sort task. The similarity-rating task was implemented in a program created by the Bureau of Glottal Affairs, Division of Head/Neck Surgery, UCLA School of Medicine. Stimuli were presented in pairs separated by

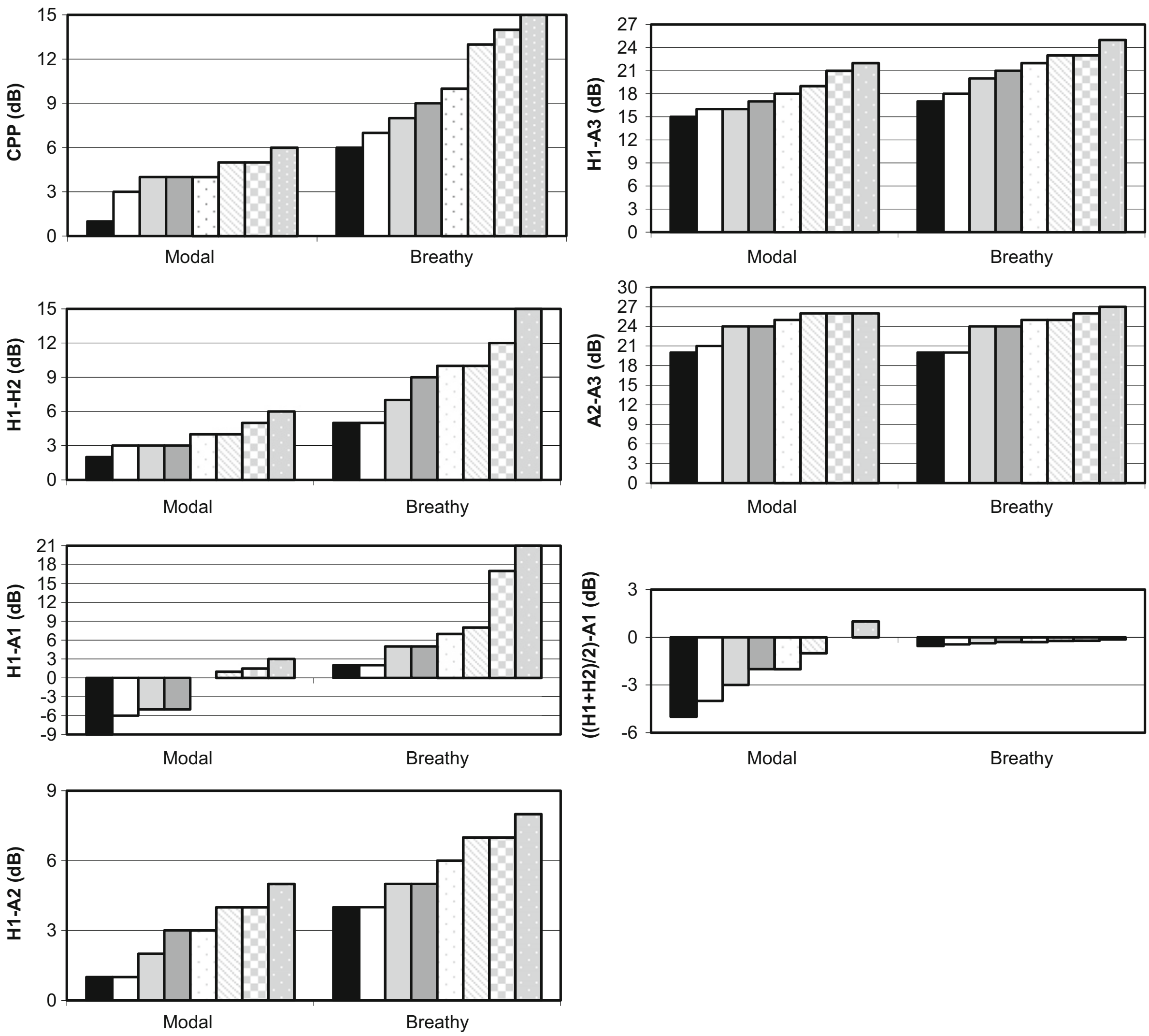


**Fig. 1.** Graph of individual CPP, H1 – H2, H1 – A1, H1 – A2, H1 – A3, A2 – A3, and ((H1+H2)/2) – A1 values for the Mazatec stimuli. In all seven graphs, a higher dB indicates increased breathiness.

0.5 s. Listeners heard both orders of the stimuli (AB and BA). The vowel that occurred first within the pair (AB or BA) was chosen randomly. (A pilot study showed that there was not a significant effect of order on the presentation of data, Esposito, 2005.)

Listeners were able to replay the pairs as often as necessary before responding, and could go back and re-listen to previous pairs. The order of the stimulus presentation was randomized across listeners. The directions were given in the listeners' native language to direct the listener to the phonation contrasts in his/her language. Listeners were asked to judge the similarity of each pair of vowels by moving an onscreen sliding cursor, where the left-most edge represented "exactly the same" and the right-most edge represented "extremely different". Listeners had fine control over the sliding cursor and could move it anywhere along the continuum. When a listener was satisfied with his/her rating of the pair, s/he could move on to the next pair by clicking the "next" button. This procedure continued until all of the pairs were rated.

The results were analyzed following the procedure used in numerous studies on the perception of pathologically disordered phonation (e.g. Kreiman & Gerratt, 1996; Kreiman, Gerratt, & Berke, 1994; Kreiman et al., 1990, to name a few). The position of the sliding cursor for each rating was converted into a numerical value, where "exactly the same" was zero and "extremely different" was 100. Within-listener agreement (that is, the extent to which listeners agreed with themselves on the perception of each pair) was determined by comparing each individual listener's response to the pair AB and his/her response to BA. If listeners were in agreement with themselves, then they should have assigned a similar rating to both AB and BA. More specifically, within-listener agreement was assessed by (1) determining the percentage of responses that differed by a value of 40 or more (roughly equivalent to the scale used in Kreiman et al., 1990) for a single listener's rating of AB and BA pairs and (2) calculating the correlation between the rating and re-rating of each pair of stimuli (that is, the rating a listener gave to the pair AB was compared to the rating that same listener gave to the pair BA).

In addition, results were analyzed using a type of multi-dimensional scaling procedure known as the individual differences model (INDSCAL), which can quantify the differences between individual listeners. INDSCAL considers the perceptual space across individual listeners and weights the importance of each dimension for each listener, producing both a group space and weights for individual subjects on each dimension. Separate INDSCAL solutions in one to five dimensions were found for each listener group. For each analysis, $r^2$ (which shows the amount of variance that is accounted for by each solution), and stress (which is an indication of how well the model fits the data) were used to find the location of the "elbow", which suggests the number of dimensions the

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002

listener groups used to make their judgments. In most cases, the $r^2$ and stress values clearly indicated the number of dimensions underlying the judgments. However, in some cases where the solution was less clear, interpretability of the solution was taken into account to determine the number of dimensions.

## 3.2. Results

### 3.2.1. Within-listener agreement

Overall, within-listener agreement was good. On average, English listeners deviated by more than a scale value of 40 on only 10.5% of the pairs and the rating and re-rating of pairs (that is, AB and BA) were correlated for the English listeners ($r^2$=0.71). On average, Spanish listeners deviated by more than a scale value of 40 on only 11.9% of the pairs and the rating and re-rating of pairs were correlated ($r^2$=0.73). Spanish listeners behaved very similarly to the English listeners, despite the allophonic breathiness in English. As expected, Gujarati listeners were more consistent in their judgments than either English or Spanish listeners. The Gujarati listeners, on average, deviated by more than a scale value of 40 on only 3.8% of the pairs and the rating and re-rating of pairs were well correlated ($r^2$=0.90).

### 3.2.2. Multi-dimensional scaling

Table 1 shows the stress and $r^2$ values for one to five dimensions for English, Spanish, and Gujarati listeners. The elbow is in bold.

For the English listeners, it was not clear whether a two- or three-dimensional solution best accounted for the variance in the data. A two-dimensional solution accounted for 48% of the variance (with a stress value of 0.45), while a three-dimensional solution accounted for 49% of the variance (with a stress value of 0.32). The third dimension did not substantially improve the solution. Moreover, when the three-dimensional solution was interpreted, it was found that the dimensions corresponded to the same acoustic properties as the two-dimensional solution.

For the Spanish listeners, stress and $r^2$ values also indicated a two-dimensional solution, which accounted for 41% of the variance in the data (with a stress value of 0.46). Additional dimensions accounted for less variance.

For the Gujarati listeners, stress and $r^2$ values indicated a one-dimensional solution, which accounted for 79% of the variance (with a stress value of 0.62). After the first dimension, the amount of variance accounted for increased gradually.

### 3.2.3. Perceptual space

The perceptual space for English, Spanish and Gujarati listeners is presented in Figs. 2, 4, and 6, respectively. Dimension one is graphed on the *x*-axis, and dimension two on the *y*-axis. For all three graphs, each point represents one of the 16 stimuli in the experiment. Voices that are perceptually similar are closer together in the space. For some graphs, it is difficult to see all 16 points because they are overlapping. For each listener group, the perceptual space was interpreted by considering which acoustic measure(s) could best account for the perceptual mapping of the stimuli. Correlation analysis was used to confirm the interpretation of the dimensions; the acoustic measure with the strongest correlation is given in each section below.

**Table 1**
Stress and $r^2$ values for one to five dimension solutions for English, Spanish and Gujarati listeners.

| Dimensions | English | | Spanish | | Gujarati | |
|---|---|---|---|---|---|---|
| | Stress | $r^2$ | Stress | $r^2$ | Stress | $r^2$ |
| 1 | 0.62 | 0.14 | 0.65 | 0.24 | **0.62** | **0.79** |
| 2 | **0.45** | **0.48** | **0.46** | **0.41** | 0.58 | 0.81 |
| 3 | 0.32 | 0.49 | 0.39 | 0.38 | 0.53 | 0.82 |
| 4 | 0.28 | 0.35 | 0.33 | 0.37 | 0.48 | 0.83 |
| 5 | 0.28 | 0.33 | 0.29 | 0.30 | 0.31 | 0.85 |

*3.2.3.1. English. The perceptual space for the English listeners*:

In Fig. 2, dimension one represents H1 – H2 ($r^2$=0.83, $p < 0.05$), with the stimuli mapped into two groups. The four left-most points (indicated by the dotted black arrow) correspond to stimuli with H1 – H2 values of approximately 10–15 dB, and the right-most points (indicated by the solid black arrow) correspond to stimuli with H1 – H2 values less than or equal to 9 dB. Dimension two represents CPP ($r^2$=0.85, $p < 0.05$), with the topmost points (indicated by the dotted gray arrow) representing stimuli with very high CPP values of approximately 13–15 dB, while the bottom-most points (indicated by the solid gray arrow) represent stimuli with lower CPP values of less than or equal to 5 dB. Intermediate points represent stimuli with intermediate CPP values.

The acoustic properties of the stimuli did not naturally lend themselves to the English listeners' perceptual mapping; that is, the stimuli were more or less continuously distributed with respect to H1 – H2 and CPP, even though the listeners divided them categorically in the perceptual space. Fig. 3 is a graph of the CPP and H1 – H2 values (in dB) of the 16 stimuli. In Fig. 3, there is no gap corresponding directly to the English listeners' division along H1 – H2. (That is, along the H1 – H2 dimension, there is not a gap between the stimulus with an H1 – H2 value of 9 dB and the stimulus with an H1 – H2 value of 10 dB.) Along the CPP dimension, there is a gap between the stimulus with a CPP

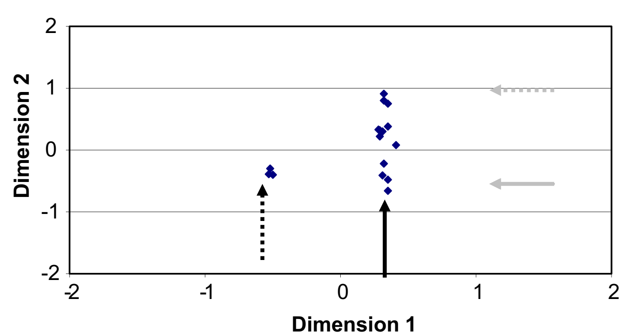


**Fig. 2.** Group perceptual space for English listeners. The scale values are arbitrary. The dotted black arrow is pointing to stimuli with H1 – H2 values of approximately 10–15 dB. The solid black arrow is pointing to stimuli with H1 – H2 values less than or equal to 9 dB. The dotted gray arrow is pointing to stimuli with CPP values of approximately 13–15 dB. The solid gray arrow is pointing to stimuli with CPP values of less than or equal to 5 dB.

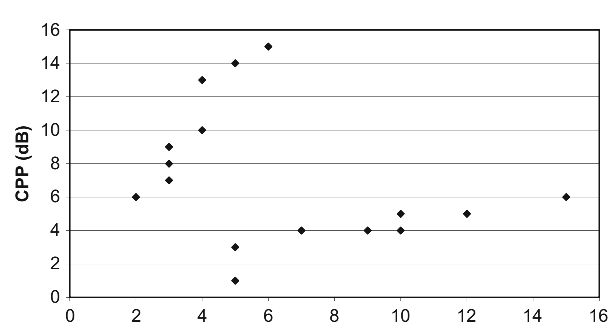


**Fig. 3.** Measured H1 – H2 values plotted against measured CPP values for the Mazatec stimuli.

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002

value of 10 dB and the stimulus with a CPP value of 15 dB, which corresponds directly to one of the English listeners' divisions along CPP in Fig. 2. However, there are other gaps in the CPP values of the stimuli which the English listeners ignored.

*3.2.3.2. Spanish.* The Spanish listeners' perceptual space is presented in Fig. 4.

Dimension one represents H1 – A1 ($r^2$=0.79, $p<0.05$), with the four left-most points (indicated by the dotted black arrow) corresponding to stimuli with H1 – A1 values of approximately −5 to −9 dB, and the two right-most points (indicated by the solid black arrow) corresponding to stimuli with H1 – A1 values of 17 – 21 dB. Intermediate points represent stimuli with intermediate H1 – A1 values. Dimension two represents H1 – H2 ($r^2$=0.84, $p<0.05$) with extreme points corresponding to extreme values. The four topmost points (indicated by the dotted gray arrow) represent stimuli with H1 – H2 values of approximately 10–15 dB, and the bottom-most points (indicated by the solid gray arrow) corresponding to stimuli with H1 – H2 values less than or equal to 9 dB.

As with the English listeners' results, the acoustic properties of the stimuli did not naturally lend themselves to the perceptual mapping presented here. Fig. 5 is a scatter plot of the measured H1 – A1 and H1 – H2 values of the stimuli. From Fig. 5, along both dimensions, there are some gaps between the stimuli. Spanish listeners' judgments corresponded to only one of the many gaps in the stimuli. Along the H1 – A1 dimension, there is a large gap between the stimulus with an H1 – A1 value of −5 dB and the stimulus with an H1 – A1 value of 0 dB. There is also a large gap between the stimulus with an H1 – A1 value of 8 dB and the stimulus with an H1 – A1 value of 17 dB. These gaps correspond directly to the Spanish listeners' divisions along H1 – A1 in Fig. 4. However, along the H1 – H2 dimension, there is no gap that corresponds directly to the Spanish listeners' division in Fig. 4.

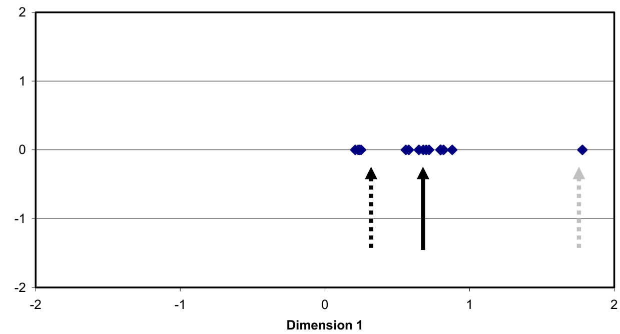


**Fig. 6.** Group perceptual space for Gujarati listeners. The scale values are arbitrary. The dotted black arrow is pointing to stimuli with H1 – H2 values of approximately 2–3 dB. The solid black arrow is pointing to stimuli with H1 – H2 values between 4 and 12 dB. The dotted gray arrow is pointing to stimuli with H1 – H2 values greater than or equal to 15 dB.

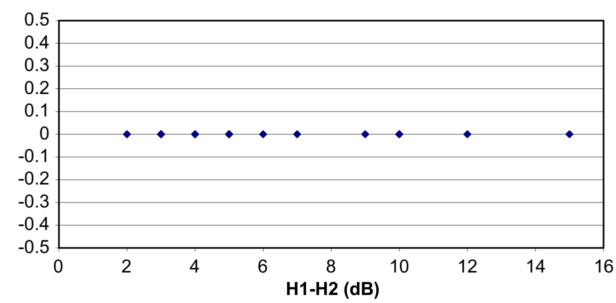


**Fig. 7.** Measured H1 – H2 values plotted against an arbitrary second "dimension" for the Mazatec stimuli.

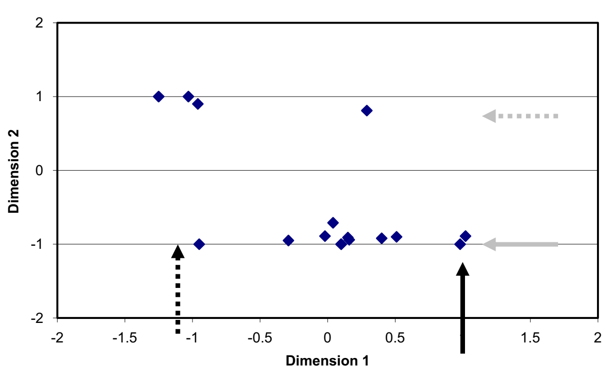


**Fig. 4.** Group perceptual space for Spanish listeners. The scale values are arbitrary. The dotted black arrow is pointing to stimuli with H1 – A1 values of approximately −5 to −9 dB. The solid black arrow is pointing to stimuli with H1 – A1 values of 17–21 dB. The dotted gray arrow is pointing to stimuli with H1 – H2 values of approximately 10–15 dB. The solid gray arrow is pointing to stimuli with H1 – H2 values less than or equal to 9 dB.

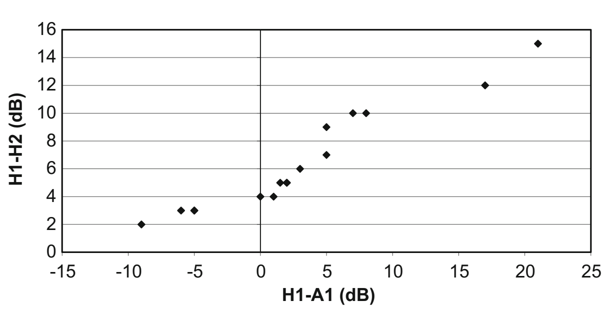


**Fig. 5.** Measured H1 – A1 values plotted against measured H1 – H2 values for the Mazatec stimuli.

*3.2.3.3. Gujarati.* For the Gujarati results (Fig. 6), there is only one dimension.

In Fig. 6, the stimuli cluster into three groups, distinguished by dimension one, which represents H1 – H2 ($r^2$=0.94, $p<0.05$). The left-most points (indicated by the dotted black arrow) correspond to stimuli with H1 – H2 values of approximately 2–3 dB, the middle set of points (indicated by the solid black arrow) correspond to stimuli with H1 – H2 values between 4 and 12 dB, and the right-most point (indicated by the dotted gray arrow) correspond to stimuli with H1 – H2 values greater than or equal to 15 dB.

Again, the acoustic properties of the stimuli did not naturally lend themselves to the perceptual mapping presented. Fig. 7 is a scatter plot of the measured H1 – H2 values of the stimuli. A second "dimension" was added to the graph along the *y*-axis to parallel Fig. 6. In Fig. 7, there are a few gaps between stimuli clustered along H1 – H2. However, most of these gaps do not correspond well to the Gujarati listeners' perceptual space (Fig. 6). One exception to this is a large gap between the stimulus with an H1 – H2 value of 12 dB and the stimulus with an H1 – H2 value of 15 dB. This corresponds directly to one of the Gujarati listeners' divisions along H1 – H2 (Fig. 6).

### *3.2.4. Individual differences*

In addition to providing a group solution, INDSCAL also indicates the degree of importance of a dimension to each individual listener by calculating a set of weights for each listener. (Only dimensions that are included in the group solution

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002

**Table 2**
Dimension weights derived from INDSCAL, showing the importance of individual dimensions to individual English, Spanish, and Gujarati listeners.

| Listeners | English listeners | | Spanish listeners | | Gujarati listeners |
|---|---|---|---|---|---|
| | Dimension 1 (H1 – H2) | Dimension 2 (CPP) | Dimension 1 (H1 – A1) | Dimension 2 (H1 – H2) | Dimension 1 (H1 – H2) |
| 1 | 0.43 | 0.32 | 0.33 | 0.29 | 0.89 |
| 2 | 0.36 | 0.33 | 0.44 | 0.42 | 0.78 |
| 3 | 0.45 | 0.33 | 0.35 | 0.24 | 0.81 |
| 4 | 0.32 | 0.25 | 0.4 | 0.31 | 0.75 |
| 5 | 0.4 | 0.35 | 0.43 | 0.41 | 0.9 |
| 6 | 0.44 | 0.42 | 0.46 | 0.42 | 0.88 |
| 7 | 0.36 | 0.26 | 0.45 | 0.34 | 0.87 |
| 8 | 0.46 | 0.34 | 0.57 | 0.37 | 0.85 |
| 9 | 0.43 | 0.35 | 0.39 | 0.34 | 0.82 |
| 10 | 0.43 | 0.34 | 0.48 | 0.36 | 0.78 |
| 11 | 0.37 | 0.31 | 0.43 | 0.36 | 0.79 |
| 12 | 0.43 | 0.32 | 0.43 | 0.35 | 0.73 |
| 13 | 0.43 | 0.35 | 0.55 | 0.4 | |
| 14 | 0.53 | 0.42 | 0.56 | 0.41 | |
| 15 | 0.34 | 0.31 | 0.61 | 0.35 | |
| 16 | 0.34 | 0.31 | 0.68 | 0.31 | |
| 17 | 0.4 | 0.31 | 0.57 | 0.25 | |
| 18 | 0.55 | 0.43 | 0.68 | 0.3 | |
| **Average** | **0.42** | **0.33** | **0.49** | **0.35** | **0.82** |

**Table 3**
Importance of individual dimensions to group English, Spanish, and Gujarati listeners.

| Language | Dimension 1 (H1 – H2) | Dimension 2 (CPP) | Dimension 3 (H1 – A1) |
|---|---|---|---|
| English | 0.39 | 0.31 | 0.08 |
| Spanish | 0.33 | 0.11 | 0.40 |
| Gujarati | 0.90 | 0.22 | 0.09 |

are accounted for in INDSCAL.) The higher the value of the weight, the more important the dimension is to that particular listener. The individual listener's weights are presented in Table 2.

For English, results showed that listeners differed in how they weighed each dimension. However, the English listeners were consistent in that all listeners agreed on the relative order of importance of the dimensions. For all 18 listeners considered individually, the first dimension of the group solution (H1 – H2) was always more important than the second dimension of the group solution (CPP).

Spanish listeners also differed in how they weighed each cue. However, as with the English listeners, the Spanish listeners agreed on the relative order of importance of the dimensions. For all 18 listeners considered individually, the first dimension (H1 – A1) was always more important than the second (H1 – H2).

For the Gujarati listeners, we cannot say anything about the relative importance of multiple dimensions beyond that fact that they only used one dimension. Therefore, individual MDS analyses were run on each Gujarati listener to see if each listener used H1 – H2 as his/her most important dimension. Results showed that the Gujarati listeners were consistent in that each listener considered individually had a one-dimensional perceptual space, and this one dimension correlated with H1 – H2 for all 12 listeners.

#### 3.2.5. *Combined-group INDSCAL*

Given that there was a common dimension across the three listener groups (that is, H1 – H2), it was possible to conduct a single INDSCAL analysis combining all the listeners, which would assign *language* rather than subject weights. Results indicated a three-dimensional group solution (stress = 0.62, $r^2$ = 0.63). A fourth dimension did not substantially improve the solution. Dimension 1 was significantly correlated with H1 – H2 ($r^2$ = 0.70, $p < 0.05$). Dimension 2 was correlated with CPP ($r^2$ = 0.56) and dimension 3 was correlated with H1 – A1 ($r^2$ = 0.52). The combined-group weights are shown in Table 3. The English listener group weighed dimension one, dimension two and dimension three, in that order. The Spanish listener group weighed dimension three as the most important dimension, followed by dimension one and then dimension two. The Gujarati group weighed dimension 1 as the most important following by dimension two and three.

## 4. Experiment 2: free-sort task

The results of the similarity-rating task established that there are in fact two perceptual categories of stimuli for all three listeners groups. In experiment 2, the free-sort task, the nature of these categories and their acoustic correlates will be examined further.

### 4.1. *Methods*

#### 4.1.1. *Listeners*

The same set of listeners used in the previous task were used here.

#### 4.1.2. *Stimuli*

*4.1.2.1. Languages and/or dialects.* Naturally produced stimuli from the following 10 languages/dialects were used: Chong (Mon-Khmer), Fuzhou (Sino-Tibetan), Green Mong (Hmong-Mien), White Hmong (Hmong-Mien), Mon (Mon-Khmer), Santa Ana del Valle Zapotec (Otomanguen), San Lucas Quiaviní Zapotec

**Table 4**
Success of the measures CPP, H1 – H2, H1 – A1, H1 – A2, H1 – A3, A2 – A3 and ((H1+H2)/2) – A1 for the 40 stimuli from Chong, Fuzhou, Green Mong, White Hmong, Mon, Santa Ana Del Valle Zapotec, San Lucas Quiaviní Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

| Languages and/or dialects | Measures | | | | | | |
|---|---|---|---|---|---|---|---|
| | CPP | H1 – H2 | H1 – A1 | H1 – A2 | H1 – A3 | A2 – A3 | ((H1+H2)/2)–A1 |
| Chong | ✓ | ✓ | – | ✓ | ✓ | – | – |
| Fuzhou | ✓ | ✓ | ✓ | ✓ | – | ✓ | ✓ |
| Green Mong | ✓ | ✓ | – | – | – | – | – |
| White Hmong | ✓ | ✓ | – | – | – | – | – |
| Mon | ✓ | – | – | ✓ | ✓ | – | ✓ |
| Santa Ana Del Valle Zapotec | ✓ | ✓ | ✓ | – | ✓ | ✓ | – |
| San Lucas Quiaviní Zapotec | ✓ | ✓ | – | – | ✓ | – | – |
| Tlacolula Zapotec | ✓ | ✓ | – | ✓ | ✓ | – | ✓ |
| Tamang | ✓ | – | – | ✓ | ✓ | ✓ | – |
| !Xóõ | ✓ | ✓ | ✓ | – | ✓ | – | – |

A check mark indicates that the measure successfully distinguished modal from breathy phonation.

(Otomanguen), Tlacolula de Matamoros Zapotec (Otomanguen), Tamang (Sino-Tibeatan), and !Xóõ (Khoisan). All of these languages/dialects possess breathy vowels[1] (whether allophonically or phonemically), and all are tonal. Recordings were either taken from the UCLA Phonetics Archive or collected by the author.

*4.1.2.2. Manipulation and measuring of stimuli.* The stimuli were composed of breathy and modal vowels excised from real words consisting of a coronal stop followed by [a]. Two breathy and two modal vowels were chosen from each language, producing a total of 40 stimuli. To control for gender, only male voices were used in the experiment. Duration was normalized to 250 ms and F0 was normalized to 115–110 Hz. The same manipulation of the stimuli performed in Section 3.1.2.1 was repeated here.

*4.1.2.3. Phonetic measurements of the stimuli.* CPP, H1 – H2, H1 – A1, H1 – A2, H1 – A3, ((H1+H2)/2 – A1, and A2 – A3 were measured following the same procedure described in Section 3.1.2.2.

*4.1.2.4. Results of the stimuli measurements.* Table 4 is a summary of the seven measures of phonation on the stimuli. (For a wider categorization of the phonations in these languages/dialects and not just on the stimuli set, see Esposito, 2006.) A check mark indicates which measures successfully distinguished breathy from modal phonation for each language; a dashed line indicates that the measure was unsuccessful. Significance could not be determined statistically because of the small number (4) of stimuli per language. Therefore, a measure was considered successful if the modal stimuli differed by at least 4 dB from the breathy stimuli on all measures except CPP. For CPP, a measure was considered successful if the modal stimuli differed by at least 1 dB from the breathy stimuli. (Pilot data showed that expert listeners could distinguish phonations that had at least a 1 dB difference in CPP, and an approximately 4 dB difference on the other measures, Esposito, 2005.)

CPP was the best measure of phonation for this stimuli set, followed by H1 – H2 (which worked in eight of the languages/dialects), H1 – A3 (a successful measure in seven of the languages/dialects), and then H1 – A2 (a successful measure in five of the languages/dialects). H1 – A1, A2 – A3 and ((H1+H2)/2) – A1 each worked in three of the languages/dialects.

Again, discriminant analysis was used to determine if these seven measures were sufficient to distinguish breathy from modal phonation across the entire sample. This was done by creating two groups, one consisting of the 20 breathy stimuli from all the languages/dialects together, and the other consisting of the 20 modal stimuli from all the languages/dialects together. The results of the discriminant analysis showed that these seven acoustic measures accounted for 91% of the variance in the data, with four measures doing most of the work: CPP accounted for most of the variance (46%), followed by H1 – H2 (27%), H1 – A2 (10%), and then H1 – A3 (8%).

*4.1.3. Procedure*

*4.1.3.1. Selecting a procedure.* While identification tasks are ideal for assessing perception of category membership, the differences between English, Spanish, and Gujarati make it impossible for all three sets of listeners to perform the same identification task. Gujarati listeners could be asked to identify the phonation they heard, but English and Spanish listeners do not have distinct categories to identify. One possible solution would be to train the English and Spanish listeners on the definitions of "modal" and "breathy", and then ask the listeners to identify the stimuli with these labels. However, training would expose listeners to the stimuli prior to the beginning of the experiment, and would also be difficult because there is no prototypical "breathy" or "modal" phonation to be trained on. Thus, the ideal task for this experiment is one that does not make a reference to a category, while still preserving some elements of an identification task. For this reason, a free-sort task was designed for this experiment.

A free-sort task is one in which subjects sort a set of items into groups or piles according to some self-chosen criterion. A type of free-sort task called a Visual Sort and Rate (VSR) was used in a study on the perception of pathologically disordered voice quality (Granqvist & Eng, 2003). In the current experiment, subjects were asked to place each stimulus (by moving an icon on a computer screen) into one of two sorting boxes,[2] based on perceived similarity. The free-sort method is advantageous because a reference to a category does not need to be made to sort the stimuli. Thus, all three listening populations can perform this task equally. An additional benefit of this method is that listeners have direct control over the presentation of the stimuli. (In tasks that allowed listeners direct control over the stimuli, listeners were more motivated to listen carefully, Logan & Pruitt, 1995.)

[1] Chong, Fuzhou, White Hmong, Green Mong, the three Zapotec languages, and !Xóõ also have creaky vowels which are not relevant to this study.

[2] Pilot data showed that subjects were confused by an unconstrained sorting task (that is, one with no boxes), and by a sorting task with three or more boxes (Esposito, 2005).

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002

*4.1.3.2. Free-sort procedure.* The free-sort task[3] was implemented in Matlab for Windows. All the stimuli were presented in one trial and arranged in a random order.

Listeners were asked to place each stimulus (by moving an icon) into one of the two sorting boxes, based on perceived similarity. Listeners were instructed to use both boxes and not to leave an item alone in a box. (This was to ensure that listeners would not sort everything into one box.)

The directions for the experiment were given in the listeners' native language. Listeners were told to sort the stimuli based on "what the voice sounds like". While these directions were vague, studies have shown that the explicitness of instruction seems to have little effect on the outcome of the task (Polka, 1992) and pilot data showed that these instructions were sufficient (Esposito, 2005).

Listeners were told to play all the stimuli at least once before sorting any of them. Listeners then picked the stimulus of their choice to sort into one of the boxes. This procedure continued until all of the stimuli were sorted. Listeners could listen to the stimuli as often as they liked, even after sorting them, and could move a stimulus out of the box if they were not satisfied with their sort. After the sorting was done, the listeners were instructed to listen to each stimulus one more time to see if they were satisfied with their sorting before concluding the experiment. This last step facilitated the comparison among the stimuli within a box.

### 4.2. Results

The results of the free-sort task were analyzed by examining how often subjects grouped a given stimulus with every other stimulus (i.e., how often each pair was put into the same sorting box). All possible pairings of the stimuli (462 pairs in all) were examined. Per-pair consistency (Section 4.2.1), correlation between acoustic properties and listeners' judgments (Section 4.2.2), and the percent of pairs that listeners correctly identified (Section 4.2.3) were calculated.

#### 4.2.1. Per-pair consistency

*4.2.1.1. Calculating per-pair consistency.* Per-pair consistency was determined by looking at how listeners' responses deviated from randomness. This measure was calculated by assessing for each pair of stimuli the percentage of listeners who placed those two samples in the same box (versus a different box), and then subtracting that value from chance (50%). The higher the absolute value, the more consistent the listeners within a language. There are 462 pairs, so there are 462 values for per-pair consistency for each group of listeners. These 462 values were then averaged across listeners within each language.

*4.2.1.2. Results.* The average per-pair consistency was 21.05% for English listeners, 20.93% for Spanish listeners, and 48.8% for Gujarati listeners. The distribution range of the per-pair consistencies for English, Spanish, and Gujarati listeners is presented in Fig. 8. The *x*-axis represents the per-pair consistency (as a percentage). The *y*-axis represents how many times listeners had *n* per-pair agreement considering all the possible pairs (462). For example, English listeners had 10% per-pair consistency on approximately 140 of the 462 possible pairs.

Results showed that there was a lack of consistency within the English and Spanish listener groups. English listeners' per-pair consistencies were in the range of approximately 10–45%, with most in the 10% range. Spanish listeners' per-pair consistencies were in the range of approximately 10–50% with most in the 10% range. Gujarati listeners' per-pair consistencies were in the range of 10–50%, with most greater than 40%. English and Spanish listeners were significantly less consistent ($p \leq 0.05$) then Gujarati listeners as determined by two one-sample *t*-tests, one comparing the English listeners' results to Gujarati results, and the other comparing the Spanish listeners' results to Gujarati results.

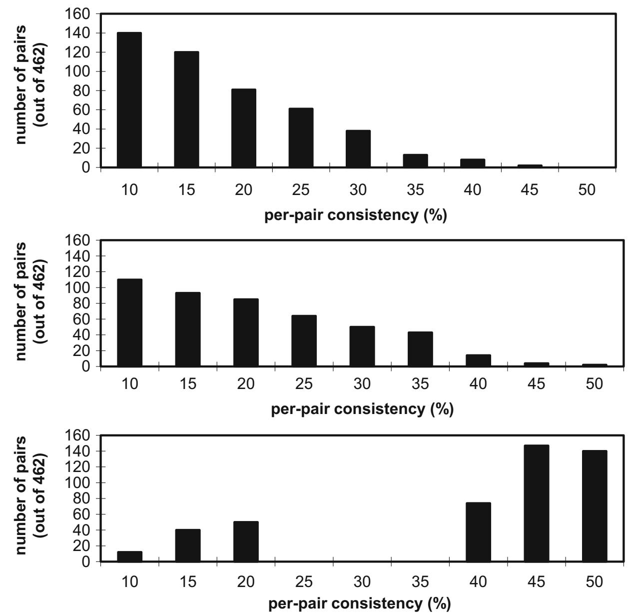


**Fig. 8.** Distribution range of per-pair consistency for English, Spanish, and Gujarati listeners.

#### 4.2.2. Correlations

*4.2.2.1. Calculating correlations.* Correlations between acoustic properties and listeners' judgments were determined for each of the seven acoustic measures based on pairs of stimuli. Listeners judgments were again coded as either putting each pair of stimuli into the same box or not. For each pair, their acoustic differences (in dB) were taken for each of the seven acoustic measures. For example, if the H1 – H2 value of stimulus A was 20 dB, and the H1 – H2 value of stimulus B was 15 dB, then the H1 – H2 difference between these two stimuli is 5 dB. The smaller the dB difference (for a given measure) between two stimuli, the more similar the stimuli, and the more likely the stimuli should be paired together. All possible pairings of stimuli along each of the acoustic dimensions (CPP, H1 – H2, H1 – A1, H1 – A2, H1 – A3, ((H1+H2)/2) – A1) and A2 – A3) were examined.

*4.2.2.2. Results.* The correlation between the measures and perceptual judgments for each listener group is presented in Table 5. A significant correlation ($p \leq 0.05$) between the acoustic measure and the perceptual judgment is marked with an asterisk.

Results for English listeners showed a significant but weak relationship between H1 – H2 and the perceptual judgments. Results for Spanish listeners showed a significant but weak relationship between H1 – H2 and H1 – A1 and listeners' judgments. There was a strong, significant relationship between H1 – H2 and Gujarati listeners' judgments.

#### 4.2.3. Percent correct

*4.2.3.1. Calculating percent correct.* Using the natural phonemic categorization of the stimuli, it was possible to determine the

[3] The script (Wilson, 2005) is available for download at www.linguistics.ucla.edu/faciliti/facilities/perception/matlab.html.

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002

**Table 5**
Correlation between the measures and perceptual judgments for English, Spanish, and Gujarati listeners.

| Measure | Listener population | | |
|---|---|---|---|
| | English $r^2$ | Spanish $r^2$ | Gujarati $r^2$ |
| CPP | 0.00 | 0.00 | 0.02 |
| H1 – H2 | 0.14* | 0.22* | 0.83* |
| H1 – A1 | 0.08 | 0.16* | 0.07 |
| H1 – A2 | 0.00 | 0.00 | 0.11 |
| H1 – A3 | 0.02 | 0.02 | 0.00 |
| A2 – A3 | 0.00 | 0.00 | 0.00 |
| ((H1+H2)/2) – A1 | 0.06 | 0.07 | 0.00 |

percent of the stimuli that listeners sorted "correctly". Language experience should hurt a listener's ability to perceive a similar contrast in a different language if the cues are different. Thus, it is likely that Gujarati listeners will not be able to correctly identify breathy and modal vowels that are not produced along H1 – H2 differences.

Percent correct was determined by counting the number of times listeners paired stimuli with the same phonation category (either breathy or modal) in the same box, divided by the total number of pairs. This was then multiplied by 100 to give a percentage. All possible pairings of the stimuli were examined.

*4.2.3.2. Results.* As expected, the Gujarati listeners had a higher percent correct (90%) than either English (54%) or Spanish (58%) listeners. Gujarati listeners incorrectly sorted only 10% of the pairs. Examination of their errors indicated that these pairs were always ones where the amplitude of H1 was low, especially when compared to H2. Most of the errors where on stimuli that came from Mon and Tamang, the two languages where breathy and modal phonation are not distinguished by H1 – H2 differences.

## 5. Summary and conclusion

Recall that the purpose of this study was to answer the question of how linguistic experience affects the perception of linguistically relevant phonation and the acoustic substrates of this perception. As expected, Gujarati listeners were more consistent in their judgments than either Spanish or English listeners, and did better at distinguishing the same type of contrast in other languages/dialects.

In terms of acoustic dimensions, in both tasks, English listeners' judgments correlated with H1 – H2. However, in the similarity-rating task, but not the free-sort task, judgments were also correlated with CPP. One possible reason for why English listeners used CPP in the similarity-rating task, but not in the free-sort task, is because the CPP values of the stimuli used in the free-sort task were very narrow (from 0.5 to 4 dB). In the free-sort task English listeners were not sensitive to small differences in noise, in that they treated stimuli with CPP values less than 4 dB as the same. But, when the range of values was greater (0–15 dB, as in the similarity-rating task), listeners' judgments were correlated with this measure, but they still treated stimuli in the range of 0–5 dB as the 'same' (i.e. they were clustered together in the perceptual mapping). Furthermore, in the similarity-rating task, English listeners consistently weighed H1 – H2 as the most important dimension after CPP. This is in contrast to Klatt and Klatt (1990) and Hillenbrand et al. (1994), who both found that noise (in the form of aspiration or CPP) was a more important cue to the perception of breathiness for English listeners than the amplitude of the first harmonic.

In both the similarity-rating task and the free-sort task, Spanish listeners based their judgments on H1 – H2 and H1 – A1. Spanish listeners were also consistent in that they weighed H1 – A1 greater than H1 – H2. When basing their judgments on H1 – H2, Spanish listeners mapped the stimuli just like the English listeners.[4] Both listener groups divided stimuli along the H1 – H2 dimension into two clusters, one with H1 – H2 values less than 9 dB, and the other with H1 – H2 values greater than 9 dB. The question arises: is 9 dB a perceptual cut-off for breathy and modal phonation? One way to address this question would be to synthesize a smooth range of stimuli with very small intervals between the stimuli (e.g., 0.1 dB), with no particular gaps. If presented with a smooth range of stimuli, would English and Spanish listeners still use the 9 dB cut-off? Furthermore, the similarities between Spanish and English listeners' judgments throughout the two experiments suggests that Spanish might have allophonic non-modal phonation of the sort found in English. Perhaps the allophonic non-modal phonation of Spanish is produced along H1 – A1 and H1 – H2, though a production study needs to be conducted to verify this idea.

Finally, in both the similarity-rating task and the free-sort task, the Gujarati listeners relied on H1 – H2. Even though Gujarati only has two phonation categories, in this experiment as well as the other two similarity-rating tasks in Esposito (2006), the Gujaratis mapped the stimuli into three clusters of approximately 2–3, 4–12 dB, and greater than or equal to 15 dB. This suggests that there is a ceiling for acceptable breathy phonation in Gujarati: anything with H1 – H2 values in the range of 2–3 dB is considered "modal", stimuli with H1 – H2 values of approximately 4–12 dB are "breathy", and stimuli with H1 – H2 values greater than or equal to 15 dB are "other". This result is similar to Bickley (1982); in her study, Gujarati listeners found vowels with H1 – H2 values of approximately 12.5 dB to be 'very breathy', vowels with H1 – H2 values of approximately 8.3 dB to be 'breathy', and vowels with H1 – H2 values of 0 dB to be 'not breathy'. Furthermore, the perceptual mapping of the stimuli reflected the breathy versus modal distinction in Gujarati to some extent. In the Gujarati sample measured for this study, the H1 – H2 values for modal vowels ranged from −1 to 4 dB, while the H1 – H2 values for breathy vowels ranged from 7 to 12 dB. There were no stimuli with H1 – H2 values higher than 12 dB in the sample measured for this study and in the sample measured by Fischer-Jørgensen (1967). However, in the present study, when forced to sort the stimuli into two groups, as in the free-sort task, Gujarati listeners placed stimuli with H1 – H2 values greater than 15 dB in the same box as stimuli with H1 – H2 values of 4–12 dB.

The perceptual mapping of the stimuli by the Gujaratis, though categorized, did not correspond to the breathy versus modal distinction in Mazatec or the language/dialects used in the free-sort task. In the similarity-rating task, the Gujaratis mapped the stimuli into three clusters, despite the fact that there are only two categories in Mazatec. In addition, in both the similarity-rating task and the free-sort task, Gujaratis sometimes judged stimuli with two different phonation categories to be perceptually similar.

In some ways, the similarity-rating task and the free-sort task are like a categorical perception task. Gujarati listeners' had sharp boundaries between stimuli, which corresponded to the boundary between breathy and modal phonation in Gujarati. Listeners did not hear stimuli with intermediate H1 – H2 values as intermediate

[4] Individual Spanish results were analyzed and it was determined that the number of years of English-language education and the number of years in the US did not play a role in listeners' performance in either the similarity-rating or free-sort task. Furthermore, the set of Spanish listeners who received extra credit for their participation did not behave differently from those who were paid for their time.

in nature, but rather always formed clusters with the stimuli. Furthermore, the Gujaratis also favored the acoustic cue associated with phonation in their native language (H1 – H2) even though other acoustic measures (such as H1 – A2 in the similarity-rating task and CPP in the free-sort task) were superior in distinguishing the phonation types of the stimuli used in the experiments.

Finally, all three listener groups' perceptual judgments correlated with H1 – H2. Cross-linguistically, H1 – H2 is one of the most common ways to produce phonation differences. H1 – H2 is a successful measure of phonation in a variety of languages (e.g., Hmong (Huffman, 1987), !Xóõ (Ladefoged & Antoñanzas-Barroso, 1985), Gujarati (Bickley, 1982), Chong (Blankenship, 1997), Jalapa Mazatec (Blankenship, 1997), allophonic variation in American English (Epstein, 1999)). Since all healthy speakers are equally capable of making all possible articulations, the reason that H1 – H2, which is correlated with the open quotient (Holmberg et al., 1995), is favored in so many languages could be a perceptual one. Perhaps, open quotient differences are naturally more salient to human listeners.

## Acknowledgements

This article is based on a larger research project guided by Bruce Hayes, Sun-Ah Jun, Patricia Keating, and Jody Kreiman. I thank them for their assistance. I would also like to thank Peter Ladefoged and Kie Zuraw who commented on earlier versions of this research.

## References

Abramson, A., Thongkum, T., & Nye, P. W. (2004). Voice register in Suai (Kuai): An analysis of perceptual and acoustic data. *Phonetica*, *61*, 147–171.
Andrade, A. (2003). *Intonational phonology of alteño Spanish*. M.A. thesis, UCLA.
Andruski, J., & Ratliff, M. (2000). Phonation types in production of phonological tone: The case of Green Mong. *Journal of the International Phonetic Association*, *30*, 37–61.
Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers, Research Laboratory of Electronics* (pp. 73–93). Boston: MIT.
Blankenship, B. (1997). *The time course of breathiness and laryngealization in vowels*. Ph.D. dissertation, UCLA.
Canfield, D. L. (1981). *Spanish pronunciation in the Americas*. Chicago: The University of Chicago Press.
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*, 423–444.
Epstein, M. (1999). *A comparison of linguistic and pathological breathiness using the LF model*. M.A. thesis, UCLA.
Epstein, M. (2002). *Voice quality and prosody in English*. Ph.D. dissertation, UCLA.
Esposito, C. M. (2003). *Phonation in Santa Ana del Valle Zapotec*. M.A. thesis, UCLA.
Esposito, C. M. (2004). Santa Ana del Valle Zapotec Phonation. *UCLA Working Papers in Phonetics*, *103*, 71–105.
Esposito, C. M. (2005). Pilot of for the effects of linguistic experience on the perception of phonation. UCLA unpublished manuscript.
Esposito, C. M. (2006). *The effects of linguistic experience on the perception of phonation*. Ph.D. dissertation, UCLA.
Fischer-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels. *Indian Linguistics*, *28*, 71–139.
Gerfen, C., & Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics*, *33*, 311–334.
Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.
Granqvist, S., & Eng, L. (2003). The Visual Sort and Rate method for perceptual evaluation in listening tests, 〈http://www.speech.kth.se/~svante/Thesis/paperI.pdf〉.
Guion, S. G., Post, M. W., & Payne, D. L. (2004). Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics*, *32*, 517–542.
Hillenbrand, J. M., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, *37*, 769–778.
Holmberg, E., Hillman, R., Perkell, J., Guiod, P., & Goldman, S. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech, Language, and Hearing Research*, *38*, 1212–1223.
Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America*, *81*, 495–504.
Kirk, P. L., Ladefoged, J., & Ladefoged, P. (1993). Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. In A. Mattina, & T. Montler (Eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson* (pp. 435–450). Missoula, MT: Universtiy of Montana Press.
Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820–857.
Kreiman, J. (1982). Perception of sentences and paragraph boundaries in natural conversation. *Journal of Phonetics*, *10*, 163–175.
Kreiman, J., & Gerratt, B. R. (1996). The perceptual structure of pathologic voice quality. *Journal of the Acoustical Society of America*, *100*(3), 1787–1795.
Kreiman, J., Gerratt, B. R., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research*, *36*, 14–20.
Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathological voice quality. *Journal of the Acoustical Society of America*, *96*(3), 1291–1301.
Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research*, *33*, 103–115.
Ladefoged, P. (1983). The linguistic use of different phonation types. In D. Bless, & J. Abbs (Eds.), *Vocal fold physiology: Comtemporary research and clinical issues* (pp. 351–360). San Diego: College-Hill Press.
Ladefoged, P., & Antoñanzas-Barroso, N. (1985). Computer measured of breathy voice quality. *UCLA Working Papers* 61 (pp. 79–86).
Lofqvist, A., & McGowan, R. S. (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics*, *20*, 93–110.
Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive nonnative phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351–378). Maryland: York Press Inc.
Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–127). Cambridge: Cambridge University Press.
Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics*, *52*, 37–52.
Silverman, D. (1997). Laryngeal complexity on Otomanguean vowels. *Phonology*, *14*, 235–262.
Silverman, D., Blankenship, B., Kirk, P., & Ladefoged, P. (1995). Phonetic structures of Jalapa Mazatec. *Anthropological Linguistics*, *37*, 70–88.
Stevens, K.N. (1988). Modes of vocal fold vibration based on a two-section model. In O. Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms, and function* (pp. 357–371). New York: Raven Press.
Stevens, K.N., & Hanson, H. (1995). Classification of glottal vibration from acoustic measurements. In O. Fujimura, & M. Hirano (Eds.), *Voice quality control* (pp. 335–342). San Diego: Singular Publishing Group, Inc.
Thongkum, T. (1987). Phonation types in Mon–Khmer languages. *UCLA Working Papers in Phonetics*, *67*, 29–48.
Upperman, G. (2004). Changing pitch with PSOLA for voice conversion 〈http://cnx.org/content/m12474/latest/〉.
Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, *31*, 181–201.
Wilson, S. (2005). Script for free-sort experiment 〈http://www.linguistics.ucla.edu/faciliti/facilities/perception/matlab.html〉.

Please cite this article as: Esposito, C. M. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* (2010), doi:10.1016/j.wocn.2010.02.002