UNIVERSITY OF CALIFORNIA

Los Angeles

The Effects of Linguistic Experience on the Perception of Phonation

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Linguistics

by

Christina Marie Esposito

2006

The dissertation of Christina Marie Esposito is approved.

Bruce Hayes

Sun-Ah Jun

Patricia A. Keating, Committee Co-Chair

Jody Kreiman, Committee Co-Chair

University of California, Los Angeles

2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VITA

February 20, 1978	Born, Islip Terrace, Long Island, New York
2000	B.A. in Anthropology and Linguistics State University of New York at Albany
2003	M.A., Linguistics University of California, Los Angeles
2001-2006	Graduate Teaching Assistant Department of Linguistics University of California, Los Angeles
2001-2006	Graduate Research Assistant Department of Linguistics University of California, Los Angeles
2004-2005	Mabel Wilson Richards Fellowship
2006 (Fall)	Instructor Department of Linguistics University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Esposito, Christina M. (2006) The effects of linguistic experience on the perception of phonation. Poster presented at the 80th meeting of the Linguistic Society of America, Albuquerque, NM.

Esposito, Christina M., Sameeruddowla Khan, and Alex Hurst (2005) Breathy nasals and /Nh/ clusters in Bengali, Hindi, and Marathi. *UCLA Working Papers in Phonetics*, no. 104.

Esposito, Christina M., Sameeruddowla Khan, and Alex Hurst (2005) Breathiness in Indic Languages. Poster presented at the 149th meeting of the Acoustical Society of America, Vancouver, BC, 17 May 2005.

Esposito, Christina M. (2005) An acoustic and electroglottographic study of phonation in Santa Ana del Valle Zapotec. Poster presented at the 79th meeting of the Linguistic Society of America, San Francisco, CA.

Esposito, Christina M. and Rebecca Scarborough (2004) An acoustic and electroglottographic study of VʔV in two indigenous American languages. Poster presented at the Meeting of the Acoustical Society of America, San Diego, CA.

Esposito, Christina M. (2004) Santa Ana del Valle Zapotec Phonation. *UCLA Working Papers in Phonetics,* no. 103.

Esposito, Christina M. (2003) The effects of f0 and position-in-utterance on phonation in Santa Ana del Valle Zapotec. *Proceedings of the XVth International Congress of Phonetic Sciences (ICPhS).*

Esposito, Christina M. (2003) The effects of f0 and position-in-utterance on phonation in Santa Ana del Valle Zapotec. Poster presented at the XVth International Congress of Phonetic Sciences, Barcelona, Spain.

ABSTRACT OF THE DISSERTATION

The Effects of Linguistic Experience on the Perception of Phonation

by

Christina Marie Esposito

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2006

Professor Patricia A. Keating, Co-Chair

Professor Jody Kreiman, Co-Chair

This study investigates the influence of linguistic experience on the perception of phonation and the acoustic properties that correlate with this perception. Listeners from Gujarati (contrasts breathy vs. modal vowels), Spanish (no breathiness) and English (allophonic breathiness) participated in a free-sort task using stimuli from various languages/dialects that differ in their phonation production, a similarity-rating task using stimuli from Mazatec, a language that contrasts phonation, and a similarity-rating task with pathologically-disordered stimuli.

It was hypothesized that 1) due to the phonation contrast in Gujarati, Gujaratis would perform more consistently than English and Spanish listeners, 2) due to the allophonic breathiness in English, English listeners would be more consistent than Spanish listeners, 3) Gujarati and English listeners' judgments would correlate with the

measure H1-H2 (amplitude of the first harmonic minus the second harmonic) because this measure correlates with their production of breathiness, and 4) Spanish listeners, in the absence of any guidance from their language, would rely on the largest auditory difference between the stimuli.

The results of the free-sort task were analyzed by determining per-pair consistency, correlations between acoustic measures and listeners' judgments, and percent of pairs sorted correctly. The results of the similarity rating tasks were analyzed using multi-dimensional scaling.

Results showed that Gujaratis did better at distinguishing breathy from modal vowels and were more consistent than either English or Spanish listeners. Despite their allophonic breathiness, English listeners did no better at distinguishing phonations than Spanish listeners did, nor were they more consistent. Gujaratis relied solely on the measure H1-H2, which is associated with the production of phonation in Gujarati. English listeners relied weakly on the measure H1-H2, which is associated with the production of phonation in English. English listeners also used cepstral peak prominence in the three similarity-rating tasks. Spanish listeners relied on H1-H2 and H1-A1 (amplitude of first formant peak).

In the similarity-rating task using pathologically-disordered stimuli, Gujaratis behaved more consistently than other listeners, who disagreed on the relevant importance of the measures that they used. All three listener types treated the pathologically-disordered stimuli in a similar way to the Mazatec stimuli.

Chapter 1: Introduction

1.1 Overview of the production and perception of phonation

Most of the phonetic research on linguistically-relevant phonation involves experimental descriptions of the production of phonation in particular languages, such as Gujarati (Fisher-Jørgensen 1967) and Mazatec (Silverman et. al 1995). Recently, this research has been expanded 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 (that is, advanced or retracted tongue root) and phonation (Guion et. al 2004), and the interaction between intonation and phonation (Epstein 2002; Esposito 2003).

However, while we now have a decent understanding of the production of linguistically-relevant phonation, little is known about its perception, especially by listeners who contrast phonation. There are only two small-scale studies on this topic; in both, Gujarati listeners were successful in their identification of breathy and modal vowels due to the phonemic contrast between these two phonation types in their language (Fischer-Jørgensen 1967; Bickley 1982).

Actually, most of what is known about the perception of phonation comes from studies where English listeners judge pathologically-disordered voice qualities, a task they often do quite poorly on with little inter-subject consistency (Kreiman et al. 1990; Kreiman et al. 1992; Kreiman et. al 1993; Kreiman and Gerratt 1996). For example, Kreiman and Gerratt (1996), in a multidimensional scaling analysis of pathologically-disordered voices, found that out of the 3160 similarity ratings made by expert listeners there was only one dimension that was judge consistently, the severity of the voice. Furthermore, Kreiman and Gerratt (2000) showed that, even in a binary classification task (breathy/not breathy), expert English listeners agreed poorly on the categorization of pathologically breathy voices. The work on the perception of pathological voices suggests that English listeners are not listening to voices in the same way, hence the inconsistency in judgments across listeners.

However, it is possible that listeners of languages that contrast phonation would be better at such tasks and would be more consistent in their judgments. Peter Ladefoged (1981: 35) once wrote, "what is a pathological voice quality in one language may be phonologically contrastive in another." There is some production evidence supporting this claim. In English, the allophonic breathiness that occurs after intervocalic /h/ is not significantly different than pathologically breathy voice (Epstein 1999). If one person's pathology is truly another person's phonemic phonation, then speakers of languages with phonation contrasts should be better at distinguishing pathologically-disordered voice qualities. Thus, one's language background could play an important role in the perception of phonation. This issue is especially interesting when we consider the fact that in a language with contrastive phonation (e.g. !Xóõ (Ladefoged et al. 1985); Zapotec (Esposito 2003)) there can be a wide variety of productions; presumably, listeners must be able to associate more than one acoustic property with a particular phonation category.

1.2 Research overview

The purpose of this dissertation is to investigate the issue of the effect of linguistic experience on the perception of phonation and the acoustic substrates of this perception, using three groups of listeners: English (which has allophonically breathy vowels), Gujarati (which contrasts breathy and modal vowels, phonemically) and Spanish (which only has modal vowels). Specifically, the following questions will be addressed:

> 1) How does linguistic experience affect listeners' perception of phonation?
> 2) Are listeners who are sensitive to linguistically-relevant phonation differences more sensitive to pathologically-disordered phonation?
> 3) What acoustic properties correlate with listeners' perception?

It is hypothesized that speakers of languages with phonation contrasts will be better at distinguishing modal from breathy phonation. Furthermore, it is predicted that these listeners will perform more consistently than those without such contrasts, both in terms of linguistically-relevant and pathologically-disordered phonations. In addition, it is hypothesized that the acoustic correlate(s) of phonation in a given language will be the most salient acoustic property for listeners of that language. These hypotheses will be discussed in further detail in Chapter 2, section 2.5.

To answer these questions, listeners will participate in a free-sort task using stimuli drawn from a wide variety of languages and/or dialects that differ in their production of phonation, in three similarity-rating tasks using stimuli drawn from Mazatec, a language that contrasts phonation, and in an additional similarity-rating task with pathologically-disordered stimuli.

Before I discuss the dissertation research in detail, I will first present some relevant background information on phonation and perception (Chapter 2). Chapter 3 will report on the free-sort task, presenting a description of the listeners and stimuli used in the study, followed by the results of the experiment. In Chapter 4, I will report on the similarity-rating tasks that used Mazatec stimuli, presenting a description of the listeners and stimuli, followed by the results, which were analyzed using multi-dimensional scaling. In Chapter 5, I will report on the similarity-rating task that used pathologically-disordered stimuli, presenting a description of the listeners and stimuli, followed by the results, which were also analyzed using multi-dimensional scaling. Chapter 6 contains a discussion and conclusion of this study, including ideas for future research.

Chapter 2: Background

2.1 Introduction

Previous research has been concerned with the production of phonation, measuring phonation, and the perception of both linguistically-relevant and pathologically-disordered phonations. These areas will be reviewed in section 2.2.

In addition, previous research on the perception of other phonetic contrasts and the acoustic cues used in speech perception will be reviewed in section 2.3 in order to help formulate hypotheses on the perception on phonation.

The phonation of English, Gujarati, and Mexican Spanish, the three languages spoken by the listeners in this study, will be discussed in section 2.4. A summary of this chapter will be presented in section 2.5.

2.2 Phonation

2.2.1 Production

Ladefoged (1971) proposed a continuum of glottal states, from most open (spread as for a voiceless sound) to most closed (a glottal stop). This continuum is reproduced in Figure 1. The three main types of phonations that occur cross-linguistically (modal, breathy, and creaky) each represent a point on the continuum.

	Most open	←	→		Most closed
Phonation type	Voiceless	Breathy	Modal	Creaky	Glottal Closure

Figure 1:Continuum of phonation types (after Ladefoged 1971), from Gordon and Ladefoged (2001).

Toward the left end of the continuum is breathy voice, which can be defined by vocal folds that vibrate, but without much contact. During breathy voice, there is minimal adductive tension and little longitudinal tension in the vocal folds, which can result in turbulent airflow through the glottis (as reviewed by Gordon and Ladefoged 2001).

The midpoint on the continuum is modal voice. Modal voice has closure that is greater than breathy, but less than creaky voice. It is characterized by vocal folds with normal adductive and longitudinal tension (as reviewed by Gordon and Ladefoged 2001).

Toward the right end of the continuum is creaky voice, which is defined by vocal folds that have the greatest degree of closure and high adductive tension, but little longitudinal tension (as reviewed by Gordon and Ladefoged 2001). In addition, creaky voice can sometimes be characterized by period doubling and/or vocal fry.

As previously mentioned, most of the work on phonation has concentrated on defining and describing breathy, modal, and/or creaky phonation in various languages. Examples include: Fischer-Jørgensen (1967) on Gujarati, Huffman (1987) on Hmong, Thongkum (1988) on Nyah Kur and Kui, Kirk et al. (1993) on Jalapa Mazatec, and, more recently, Andruski and Ratliff (2000) on Green (H)Mong, Wayland and Jongman (2003) on Khmer, Abramson et al. (2004) on Suai, and Esposito (2003 and 2004) on Zapotec, to name a few.

These and other similar studies have demonstrated a number of cross-linguistic similarities in the production of phonation. For example, it has been shown that Hmong (Huffman 1987), Mazatec (Blankenship 1997), and Green (H)Mong (Andruski and Ratliff 2000) all distinguish breathy vowels from modal ones by H1-H2 differences interpreted as reflecting differences in open quotient (the portion of time the glottis is open during each glottal cycle).

However, a number of differences across languages have also been revealed. For example, in Santa Ana del Valle Zapotec (hereafter, SADV Zapotec), for male speakers, breathy phonation is distinguished from modal phonation by H1-A3 (assumed to reflect the speed of closure) and not by H1-H2 (Esposito 2004).

Even within a language, there can be differences in the production of phonation. While male SADV Zapotec speakers distinguish phonation categories along H1-A3, female speakers use a different acoustic dimension (H1-H2) (Esposito 2004). This acoustic difference suggests a difference in production, which was confirmed in an electro-glottographic study (Esposito 2005); the SADV Zapotec men and women were using their vocal folds differently to produce the same linguistic contrast. Thus, listeners hear different acoustic patterns for a phonation category, and must weigh the available acoustic cues differently for different speakers.

2.2.2 Pathologically-disordered phonations

The linguistically-relevant phonations described above are not all that different from the phonations produced by some speakers with vocal fold disorders. (However, it is important to note that the linguistically-relevant phonations are produced by speakers with healthy, controllable vocal folds, while pathologically-disordered phonations are the result of uncontrollable physical problems.) As Ladefoged (1983) wrote, "... one person's voice disorder is another person's phoneme". Furthermore, both pathological and linguistic breathiness are described as being produced by vocal folds that vibrate but without much contact (Laver 1980; Ladefoged and Maddieson 1996). Epstein (1999) provided evidence for the similarity between linguistically-relevant phonations and pathologically-disordered ones; her results showed that, in English, the allophonic breathy voice that occurs after intervocalic /h/ is not significantly different from pathologically breathy voice.

2.2.3 Measuring phonation

There are numerous acoustic properties that can be useful measures of the linguistically-relevant and pathologically-disordered phonation discussed previously. In this section, I will briefly discuss some of the major methods for measuring non-modal phonation from an audio signal. (Measures based on the glottal source will not be discussed.)

2.2.3.1 Periodicity

One measure of aperiodicity is cepstral peak prominence (CPP). A cepstrum is an inverse spectrum created by taking the Fast Fourier Transform (FFT) values of a spectrum. A highly periodic signal should show well-defined harmonics and therefore, have a higher cepstral peak. Non-modal phonation has less distinct harmonics and, therefore, a lower cepstral peak. Cepstral peak prominence has been a reliable measure of the aperiodicty associated with non-modal phonation not only in English (Hillenbrand et al. 1994), but also in Mazatec and Chong (Blankenship 1997).

2.2.3.2 Spectral measures

There are many different spectral measures that can be taken from an FFT. Figure 2 is an FFT of a modal vowel with H1, H2, H4, A1, A2, A3, and A4 labeled. H1 refers to the amplitude of the first harmonic, and Hn to the amplitude of the nth harmonic. A1 refers to the amplitude of the first formant peak, and An to the amplitude of the nth formant peak.

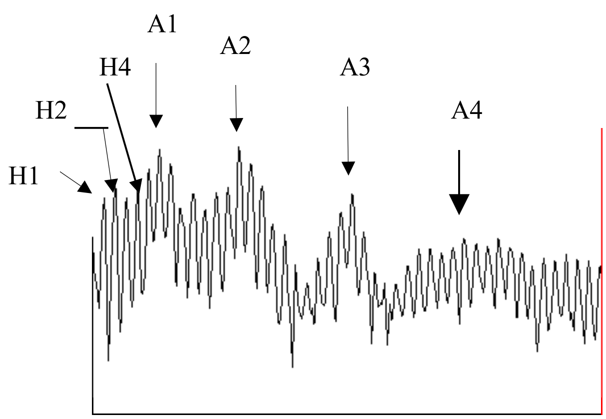


Figure 2: FFT of a modal vowel with H1, H2, H4, A1, A2, A3, and A4 labeled.

Primarily, the difference between the amplitudes of the first and second harmonics (H1-H2), which correlates with the portion of the glottal cycle in which the glottis is open (the open quotient), has been used to distinguish modal and breathy phonation (e.g. Gujarati (Fischer-Jørgenson 1967; Bickley 1982)). Recently, another low frequency measure, H2-H4 (Kreiman et al. 2006), has also been used to measure phonation, but in pathologically-disordered voices. Other studies have made use of the relationship between H1 and harmonics exciting higher formants (for example, A1, A2, A3, and A4), which correlates with the abruptness of vocal fold closure. These include: H1-A3 (Stevens and Hanson 1995; Blankenship 1997), H1-A1 or H1-A2 (Ladefoged 1983) and the average of H1-H2 compared to A1 (Stevens 1988). Other studies have used the relationship of higher formants to lower ones, such as A2-A3 (Klatt and Klatt 1990).

The various spectral measures have been associated with physiological characteristics. For example, Holmberg et al. (1995) showed that the difference between H1 and H2 correlated with the portion of time the glottis was open during each glottal cycle (the open quotient). When the vibration of the vocal folds has a large open quotient, the fundamental frequency (H1) dominates the spectrum. Thus, for breathy phonation, which has a large open quotient, the spectrum is dominated by H1 and the measure of H1-H2 is largely positive for breathy phonation.

Stevens (1977) suggested that measures of spectral slope (that is, a measure which reflect the amplitude difference between the fundamental and higher harmonics) correlated with the abruptness of vocal fold closure. When the vocal folds come together abruptly, it excites the higher frequencies of the vowel. For breathy phonation, which is characterized by vocal folds that close slowly, the higher harmonics are lower than the fundamental frequency. Measures of the spectral slope are largely positive for breathy phonation.

Because they reflect different aspects of phonation, the different measures of spectral tilt do not always distinguish phonation types, even within a single language. For example, Blankenship (1997) found that in Mpi, H1-H2 was a more reliable measure of phonation on vowels with a high tone than with a mid or low tone. This suggests two possibilities: 1) that in Mpi, the phonations of high tone vowels differ in their open quotient, while the phonations of low tone vowels do not, 2) that the measure itself is frequency-sensitive. (These two possibilities could be tested by comparing speakers with very different f0 ranges.)

There are some limitations to spectral measures. Spectral measures that are dependent on the lowest harmonics are most accurate on low vowels, because the frequency of the first formant affects the level of the harmonics. Spectral measures are also sensitive to the amount of damping of the formants and are not always accurate for breathy vowels (because there is increased damping due to the lack of complete glottal closure) (Ní Chasaide and Gobl 1997). The last limitation of spectral measures is their sensitivity to changes in f0. Any shift in the relationship between F1 and f0 can result in an inaccurate measure. However, spectral measures have successfully distinguished phonation in many languages (e.g. Gujarati (Fisher-Jørgensen 1967); Chong and Mazatec (Blankenship 1997), Zapotec (Esposito 2003), etc.).

2.2.4 Perception of linguistically-relevant phonation

There have been two small-scale studies (Fisher-Jørgensen 1967; Bickley 1982) on the perception of phonation by listeners of a language with contrastive phonation. Fisher-Jørgensen (1967) conducted a perception experiment with Gujarati listeners using naturally produced Gujarati stimuli. She 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, when presented in a random order. She found that, while listeners had difficulty in this task, some were able to successfully identify breathy vowels, especially in cases where the spectrum was dominated by the fundamental frequency (H1). Other cues, such as duration and f0, had little importance in the identification of breathy vowels. From this Fisher-Jørgensen concluded that the amplitude of H1 was the most salient cue to breathy vowels for Gujarati listeners.

Bickley (1982) also examined the perceptual correlates of breathiness by obtaining judgments from Gujarati listeners. Prior to a perception task, she asked two expert listeners (one a Gujarati speaker, the other an English speaker) to judge ten !Xóõ vowels for degree of breathiness. The listeners agreed that the breathiest vowels were the ones with the greatest difference between the first and second harmonic. This motivated the synthesis of a continuum of breathy voice in which the amplitude of the first harmonic and the degree of aspiration noise were varied. (It is not clear why aspiration noise was varied, except that it correlates with breathiness for some !Xóõ speakers.) The synthesized vowels were concatenated with Gujarati consonants to form real Gujarati words. These were played to Gujarati listeners, who were asked to identify the word they heard. Results showed that vowels with the highest amplitude of the first harmonic were consistently judged to be breathy by the Gujaratis. More specifically, Gujarati listeners found vowels with H1-H2 values of approximately 12.5 dB to be 'very breathy', vowels with an average H1-H2 value 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 average H1-H2 value of 0 dB to be 'not breathy'.

Currently, these two studies taken together are our best understanding of the perception of phonation by Gujaratis. These studies indicate that a dominant first harmonic will be the most salient acoustic cue to breathy phonation for Gujaratis.

2.2.5 Perception of non-linguistically-relevant phonations

Most of what is known about the perception of phonation comes from studies where English listeners judge pathologically-disordered voice qualities, a task they often do quite poorly on, with little inter-subject consistency (e.g. Kreiman et al. 1990; Kreiman et al. 1992; Kreiman et al. 1993; Kreiman and Gerratt 1996, etc.). For example, Kreiman and Gerratt (1996), in a multidimensional scaling analysis of judgments of pathologically-disordered voices, found that out of the 3160 similarity ratings made by expert listeners, there was only one dimension that was judged consistently, the severity of the perceived pathology. Furthermore, Kreiman and Gerratt (2000) showed that, even in a binary classification task (breathy/not breathy), expert English listeners agreed poorly on the categorization of pathologically breathy voices. In fact, listeners were not even able to agree in their judgments of how breathy a voice was (Kreiman and Gerratt 1999), or whether a voice should even be labeled as breathy (Kreiman and Gerratt 2000). The work on the perception of pathologically-disordered voice qualities suggests that listeners are not judging voices in the same way, hence the inconsistency across listeners.

An additional study, Kreiman et al. (2006), attempted to predict the acoustic cues to the perception of pathologically-disordered phonation. This study used Principle Components Analysis to test 78 measures (of the source or of the signal) for the amount of acoustic variance each individual measure accounted for in a sample of pathological voices. Of the 19 measures taken from the speech signals, four were relatively independent and contributed significantly to accounting for the variance: H1-H2, the slope of N harmonics ($3 \leq N \leq 9$), the slope of all harmonics in the spectrum, and H2-H4. By hypothesis, these dimensions are most likely to be informative to perceivers trying to distinguish the voice qualities in the sample. However, this hypothesis has not been tested.

Other studies have examined the perception of phonation which is neither linguistically-relevant nor pathologically-disordered. Klatt and Klatt (1990), who used synthesized stimuli, found that aspiration noise, rather than increased amplitude of the fundamental, was perceptually the most important cue to voice quality variations for English listeners. Furthermore, Hillenbrand et al. (1994), who studied 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. In addition, Shrivastav and Sapienza (2006), using synthesized stimuli, determined the smallest change in signal-to-noise ratio that would result in a change in breathiness. Results showed that listeners needed as much as a 20 dB increase in aspiration noise to perceive a change in breathiness against a modal voice, and an 11 dB increase in aspiration noise to perceive a change in breathiness against a severely breathy voice.

2.2.6 Perception of non-spectral cues in languages with phonation contrasts

In addition, there are two studies, Abramson et al. (2004) and Gerfen and Baker (2005), on the perception of non-spectral cues (f0 and amplitude) in languages with phonation contrasts. Abramson et al. (2004) studied the perception of phonation in the

Kuai dialect of Suai, a language that distinguishes between a modal and a breathy register, using both natural and synthesized stimuli. Unfortunately, Kuai turned out not to be ideal for studying phonation. Not all of the Kuai listeners perceived the differences among the natural tokens; listeners appeared to use f0, rather than voicing source differences, in their categorization.

Gerfen and Baker (2005) showed that, in production, Mixtec laryngealized vowels are distinguished from modal ones by four acoustic phonetic properties: a drop in 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 drops in F0 or drops in amplitude could be used by listeners to identify phonation contrasts in the absence of spectral cues.

2.3 Studies on the perception of other phonetic contrasts

It is expected that the perception of phonation will follow the general principles of within- and cross-language perception. Therefore, in the next section, we will review research on other areas of speech perception.

2.3.1 Perception of native phonemes

Early work on speech perception showed that listeners perceive speech sounds based on native language categories with sharp boundaries between perceptual categories. While changes in the stimuli might be gradual, listeners do not perceive them this way. Instead, they are perceived as categories based on the listener's native language. Many contrasts have been shown to be perceived categorically, such as voice onset time (VOT) and consonant place of articulation. For example, Liberman et al.

(1957) synthesized formant transitions to produce a continuum from /b/ to /g/ and found that, in an identification task, English listeners' responses showed sharp boundaries between stimuli. Subjects did not hear stimuli with intermediate formant values as intermediate in nature. Instead, they were perceived to be belonging to the place categories of English. In a discrimination task, listeners performed poorly on stimuli drawn from a single category along the continuum, but there were clear discrimination peaks at boundaries between categories.

In many of these early studies on perception, all but the tested cue were neutralized in the stimuli. However, in real languages there are multiple acoustic cues. It was only later that researchers looked at more realistic multi-dimensional variation (e.g. Terbeek 1977, Forrest et al. 1988, Iverson and Kuhl 1996, etc.). For example, Terbeek (1977), in a multi-dimensional scaling study of vowel perception by English, German, Thai, Turkish, and Swedish listeners, used naturally produced vowel stimuli, which differed in multiple dimensions (such as formant frequencies, amplitude, etc.) and found that perceptual judgments differed significantly based on the listeners' language background.

From the studies discussed in this section, we can see that listeners are bound by their own language categories, and the basis of their judgments is the acoustic correlates of those categories in their language. An experiment on the perception of phonation should yield similar results. That is, English, Gujarati, and Spanish listeners will be bound by the phonation categories (or lack of categories) in their native language. Furthermore, their judgments should correlate with the production of phonation in their native language. For example, Gujarati listeners' judgments should correlate with H1-H2 because this reflects the production of phonation in Gujarati. (The predicted acoustic correlates to English and Spanish listeners' perception will be presented (in section 2.5) after first discussing the phonations of these two languages in greater detail.)

2.3.2 Perception of non-native speech sounds

Cross-language speech perception studies have shown that linguistic experience can limit listeners' sensitivity to speech sounds that are not part of their native language inventory of contrasts. A well-known example is the perception of English /ɹ/ and /l/ by Japanese listeners (Strange and Dittmann 1984; Ingram and Park 1998; Yamada and Tohkura 1992a, 1992b, to name a few). In Japanese, there are no native phonemes corresponding to English /ɹ/ or /l/; the single Japanese liquid is a tap. It is difficult for Japanese listeners to perceive differences between these sounds, demonstrating the importance of native language inventory for perception. Japanese listeners' performance on the perception of English / ɹ / and /l/ is analogous to Spanish listeners' hypothesized performance on the perception of modal and breathy vowels. It is expected that Spanish listeners will perceive phonation differences poorly, because they do not have a category from their native language to base their perception on.

However, phonemic inventories cannot always explain perceptual findings. Perception of non-native speech sounds has been shown to vary based on the contrast being studied. For example, listeners are able to perceive differences between non-phonemic stimuli if their language uses a similar phonetic/allophonic contrast. Werker et al. (1981) found that English listeners were better at distinguishing Hindi aspirated /t^h/

from breathy aspirated /d^{ɦ}/ than they were at distinguishing dental /t̪/ from retroflex /ʈ/. One possible explanation for this finding is that English listeners have relevant native language experience with similar VOT contrasts (e.g. /t/ and /d/), but not with dental versus retroflex contrasts.

The dissimilarity of stimuli to any native contrast has also been shown to enhance performance. Best et al. (1988) argued that American English listeners successfully discriminated Zulu clicks because they are unlike any speech sound found in English. Thus, while native language inventory greatly affects perception, the ability to perceive a contrast cannot be predicted from native language inventory alone. Therefore, it is possible that Spanish listeners will be able to perceive some phonation differences because phonation is dissimilar to any native contrast in Spanish.

2.3.3 Perception of allophones

Other studies have examined the extent to which allophonic representations play a role in perception. Beddor and Strange (1982) showed that English listeners are sensitive to vowel nasalization that occurs allophonically before a nasal consonant and could successfully identify oral and nasal vowels, even though this is not a phonemic contrast in English. However, they required more nasalization (in the form of greater velar port opening in the articulatory synthesizer) to identify vowels as nasal than did Hindi listeners, who have a phonemic oral-nasal vowel contrast. From this study, it is hypothesized that English listeners will be sensitive to allophonic breathiness, but might require a greater degree of breathiness (i.e. a larger H1-H2 value) to identify breathy vowels than Gujarati listeners who contrast breathy and modal vowels.

2.3.4 Language-specific acoustic cues to perception

Several experiments have demonstrated the importance of specific acoustic cues to listeners in perception tasks (e.g. Gandour and Harshman 1978, Strange et al. 1979, etc.). Listeners place importance on different acoustic cues depending on their language background. For example, Fox et al. (1995), in a multi-dimensional scaling (MDS) analysis of the perception of vowels by Spanish and English listeners, found three dimensions underlying the English listeners' responses (roughly vowel height, vowel frontness, and centrality) and two underlying the Spanish listeners' responses (vowel height, and frontness). Each of these dimensions correlated with an acoustic measure. Vowel height correlated with F1 and F1-F0; frontness correlated with F2, and F2-F1; centrality correlated with F3-F2 and the Euclidean distances between the vowels in the perceptual space. It is believed that English listeners used more dimensions (and therefore more acoustic properties) because English has a larger vowel inventory than Spanish.

While most studies demonstrate that the native language of the listener determines the relative importance of acoustic cues, this is not always the case. For example, Bohn and Flege (1990), in their study on the perception of English vowels by Mandarin and Spanish listeners, found that listeners identifying English vowels relied more heavily on durational cues, than spectral ones. This is a surprising result because neither Spanish nor Mandarin uses duration to distinguish vowels. The authors hypothesized that Mandarin listeners might be sensitive to durational cues due to the duration differences among Mandarin tones. Spanish listeners, on the other hand, might have used durational cues since they could not use spectral ones to differentiate the English vowels that did not exist in Spanish. Thus, they were forced to rely on a cue they had no native language experience with.

Based on the studies mentioned above, we expect English, and Gujarati listeners to base their judgments on the relevant cues in their native language. For Spanish listeners in particular, listeners will have to rely on a cue that they have no native language experience with.

2.4 Phonation in English, Gujarati, and Spanish

In this section, I will discuss the three languages of the listeners for this study: English, Gujarati, and Spanish. Specifically, the occurrence of creaky and/or breathy voice, in addition to modal voice, will be discussed.

2.4.1 English

Phonemically, English only has modally phonated vowels. However, there are some cases where phonemically modal vowels can be produced with allophonic creaky or breathy voice. Examples of allophonic non-modal phonation in English are: 1) creakiness at the beginning of vowel-initial words associated with an allophonic glottal stop (Dilley et al. 1996), 2) creakiness associated with the ends of sentences and paragraphs (Kreiman 1982), 3) non-modal phonation at phrase boundaries (Pierrhumbert and Talkin 1992; Epstein 2002) and focal pitch-accented words (Epstein 2002), and 4) breathiness on vowels after /h/ (Löfqvist and McGowan 1992, Ladefoged 1983; Epstein 1999). Epstein (1999) showed that the allophonic breathiness in English is produced by difference in the open quotient. Due to the allophonic breathiness in English, listeners might be sensitive to phonation contrasts, though, probably, not as sensitive as Gujarati listeners, whose native language has a phonemic phonation contrast.

2.4.2 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). Historically, the breathy vowels developed from [vowel + h + vowel] sequences, [h + vowel] sequences, and from the blending of the aspiration of a final voiced aspirated stop with the preceding vowel (Fisher-Jørgensen 1967).

Fisher-Jørgensen (1967) found that Gujarati breathy vowels are characterized by greater airflow than modal ones. There is 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 is a dominant fundamental frequency (H1); on average, H1 was 4.37 dB higher for breathy vowels than for modal ones (across six speakers).

Fisher-Jørgensen (1967) also found that modal and breathy vowels can be produced with a wide variety of intonational contours (i.e. rising, rising-level, rising-falling, falling, falling-rising, or a level tone).

To confirm some of Fisher-Jørgensen's results about the height of H1 in the speech of present-day Gujarati speakers residing in California, 6 Gujarati speakers were recorded producing 10 mono-syllabic Gujarati words with the vowel [a]. Five of the 10 words had a breathy vowel, and the other five had a modal vowel. Figure 3 represents the averages of H1-H2 (in dB) graphed on the y-axis.

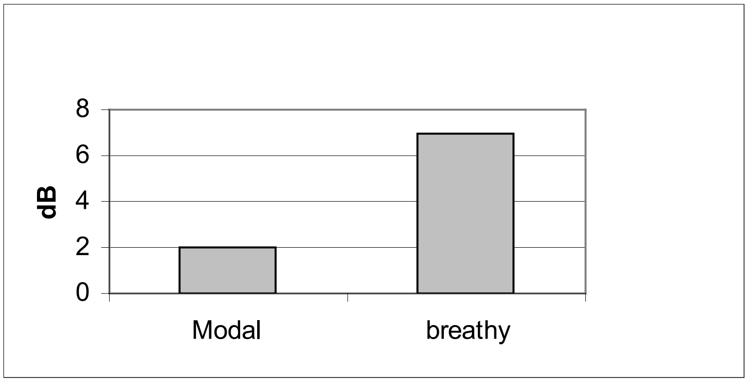


Figure 3: Average H1-H2 for Gujarati modal and breathy vowels.

From this graph, we can see that breathy vowels in Gujarati, on average, do have a higher H1-H2 value (and therefore a dominant H1) than modal vowels.

2.4.3 Mexican Spanish

Like English, Mexican Spanish does not have contrastive phonation, and phonemically has only modally phonated vowels. 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, ten native speakers of Mexican Spanish were recorded producing ten 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 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, 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). Therefore, these dialects were avoided.

2.5 Summary

As previously mentioned, this investigation centers on the following questions:

1) How does linguistic experience affect listeners' perception of phonation?
2) If listeners are sensitive to linguistic phonation differences, are they also sensitive to pathologically-disordered voice qualities?
3) What acoustic properties correlate with listeners' perception?

Previous literature on phonation and speech perception provides insight into the potential results of this study. Cross-language speech perception studies have shown that linguistic experience can limit listeners' sensitivity to speech sounds that are not part of their native language inventory (recall the results on the perception of English /ɹ/ and /l/ by Japanese listeners (Strange and Dittmann 1984; Ingram and Park 1998; Yamada and Tohkura 1992a, 1992b, to name a few). It is hypothesized that Gujarati listeners, who have a contrast between breathy and modal phonation in their native language, will be successful in distinguishing modal and breathy vowels, especially when the phonation contrasts are produced along H1-H2 difference, with uniform judgments across subjects. For any language where the phonation differences are not produced along H1-H2, the Gujaratis should be no better than the English and Spanish listeners, because Gujaratis do not have any native language experience with phonations produced along dimensions other than H1-H2.

In addition, it is predicted that Spanish listeners will perceive phonation differences poorly because they do not have a category from their native language on which to base their perceptions. However, while native language inventory greatly affects perception, the ability to perceive a contrast cannot always be determined by inventory alone. Studies have shown that the dissimilarity of stimuli compared to native contrasts can enhance performance (e.g. Best et al. 1988). Thus, it is possible that Spanish listeners will be able to perceive some phonation differences.

For English, the picture is less clear. Studies on the perception of allophonic variation (such as Beddor and Strange (1982)) suggest English listeners might be sensitive to breathy phonation because it occurs allophonically in English. However, results of studies on the perception of pathologically-disordered phonation suggest that English listeners will distinguish phonations poorly (e.g. Kreiman et al. 1990; Kreiman et al. 1992; Kreiman et al. 1993; Kreiman and Gerratt 1996, etc.).

There has not been any work done on the perception of pathologically-disordered phonations using listeners of languages with phonation contrasts. However, we know from previous research that at least some pathologically-disordered phonations resemble linguistically-relevant phonations (e.g. Ladefoged 1983; Epstein 1999). Therefore, one might hypothesize that Gujaratis will be more consistent in their judgments of pathologically-disordered phonations based on the breathy phonation category in their own language, while Spanish listeners will lack inter-subject consistency due to their lack of a breathy phonation category. For English listeners, previous research on the perception of pathologically-disordered phonation (e.g. Kreiman et al. 1990; Kreiman et al. 1992; Kreiman et al. 1993; Kreiman and Gerratt 1996, etc.) has already shown that English listeners lack inter-subject consistency.

The last question to address is which cues listeners will attend to when listening to phonation contrasts. We known from previous work discussed in section 2.3.4 that listeners place importance on different acoustic cues depending on their language background. Based on Fisher-Jørgensen (1967) and Bickley (1982) it is predicted that H1-H2 will be the most salient acoustic property to Gujaratis because this correlates with their production of breathy vowels (a claim that will be further verified in the course of the present study). For English listeners, it is also believed that they will use H1-H2 differences because this correlates with their production of allophonic breathiness (Epstein 1999). In addition, based on Klatt and Klatt (1990) and Hillenbrand et al. (1994) it is believed that English listeners will also use changes in noise to make their judgments. For Spanish listeners, their lack of a phonation contrast makes it difficult to make a prediction about what cue(s) they might try to use. But, in the absence of any guidance from their language, it is expected that Spanish listeners will rely on the largest auditory difference within the stimuli set.

Chapter 3: Free-Sort Experiment

3.1 Introduction

This investigation reported on in this chapter centers on the following questions:

1) How does linguistic experience affect listeners' perception of phonation?
2) What acoustic properties correlate with listeners' perception?

It is hypothesized that Gujarati listeners, who have a contrast between breathy and modal phonation in their native language, will be successful in distinguishing modal from breathy vowels with uniform judgments across subjects. Spanish listeners, who do not have a breathy category from their native language on which to base their perceptions, will perceive phonation differences poorly and have little inter-subject consistency. For English listeners, the picture is less clear. As previously mentioned, studies on the perception of allophonic variation suggest that English listeners might be sensitive to breathy phonation because it occurs allophonically in English. However, results of studies on the perception of pathologically-disordered phonation suggest that English listeners will distinguish phonations poorly and have poor inter-subject consistency.

Furthermore, it is predicted that H1-H2 will be the most salient acoustic property to Gujarati and English listeners because this correlates with their production of breathy vowels. In addition, it is believed that English listeners will also use changes in noise to make their judgments. For Spanish listeners, it is predicted that they will rely on the largest auditory difference between the stimuli because they do not have any guidance from their native language.

To answer these questions, an experiment involving naturally produced breathy and modal vowels from a variety of languages/dialects was carried out. Gujarati, English, and Spanish listeners participated in a free-sort task, where the stimuli were sorted into groups based on a self-chosen criterion.

The organization of this chapter is as follows: section 3.2 will explain the methods used in this study, including a description of the listeners (section 3.2.1) and the stimuli (section 3.2.2.), section 3.3 will explain the general procedure for the free-sort task. section 3.4 will be a presentation of the results, and section 3.5 will summarize the chapter.

3.2 Methods

3.2.1 Listeners

3.2.1.1 English

Eighteen native speakers of American English participated in this experiment. Of the eighteen listeners, eight were male; ten were female. None of the listeners had any experience with languages with non-modally phonated vowels (except for the allophonic non-modal phonation found in English), nor were any of the listeners familiar with the purpose of the study. All of the listeners were students at the University of California, Los Angeles, and all received extra credit in a linguistics course for their participation.

3.2.1.2 Gujarati

Thirteen native speakers of Gujarati participated in this experiment. Of the thirteen listeners, six were male; seven were female. All of the listeners were born in Gujarat, India, but now live in the U.S. and all continue to speak Gujarati on a daily basis.

All are bilingual, and speak Indian English natively in addition to Gujarati. Each subject received $15 for his/her participation.

3.2.1.2.1 Screening task: can Gujaratis perceive phonation differences in their own language?

According to one researcher "there is now a tendency [in Gujarati] to pronounce the original murmured vowels without any murmur quality" (Nair, 1979:22). (Recall that some of studies Fisher-Jørgensen's (1967) Gujarati listeners had difficulty perceiving a difference between modal and breathy phonation, so it is possible that some speakers are merging the two categories.) If such a tendency does exist, then it is possible that Gujarati speakers would not perform any differently than English or Spanish listeners in perceiving phonation differences. (Note, however, that the ability to perceive a sound does not depend on the ability to produce it. Tees and Werker (1984) found that English speakers who had experience with Hindi very early in life, but had not learned to speak the language, were able to perceive voicing contrasts that are phonemic in Hindi, but absent in English. Thus, it is possible that the Gujaratis who no longer produce a phonation contrast, might still be able to perceive one, due to early childhood exposure to the contrast from other, more conservative speakers.) It was therefore considered necessary to screen the Gujarati subjects, to determine if they could perceive the differences between the breathy and modal vowels in their own language. In the present experiment, perception of Gujarati by Gujaratis will only be studied in this screening task.

3.2.1.2.1.1 Screening task

Each Gujarati listener participated in a screening task. The screening task was randomly ordered with respect to the other tasks in the experiment.

3.2.1.2.1.1.1 Method

3.2.1.2.1.1.1.1 Stimuli

Each Gujarati listener listened to 10 naturally produced, whole-word, Gujarati minimal pairs. Seven of the pairs differed only in phonation. The other three pairs were fillers, and differed in the place of articulation of one consonant. The stimuli, drawn from Gujarati recordings from the UCLA Phonetic Archive, included common words that were familiar to all the listeners.

3.2.1.2.1.1.1.2 Procedure

The stimuli were presented as icons on a computer screen using Microsoft's PowerPoint. Listeners were instructed to click on the icons to listen to each stimulus and to identify the word they heard by circling the best English translation(s) of the word on a form provided to them. Listeners were told that a word could have one meaning, or multiple meanings, and were given an example of English homonyms. This way, listeners who merged breathy and modal vowels were free to identify a word with modal phonation, such as [kan] "ear" as both "ear" and "Krishna" (which would be [ka̤n] to a more conservative speaker). (An alternative method would have been to have the listeners indicate the word from two choices written in Gujarati. However, this is problematic because many Gujarati speakers living in the US are not literate in Gujarati.)

3.2.1.2.1.1.2 Results of the screening task

All but one of the Gujarati listeners were able to correctly identify the words in question. The listener who was not able to correctly identify the words also had difficulty producing spontaneous sentences in Gujarati (the spontaneous sentence task will be discussed in section 3.3). This listener was eliminated from the study due to her great degree of language loss. Thus, the results presented below are for 12 Gujarati listeners, not 13.

3.2.1.3 Mexican Spanish

Eighteen native speakers of Mexican Spanish participated in this experiment. Of the eighteen listeners, 7 were male; 11 were female. All of the subjects were born in Mexico, but are now living in the US, and all spoke English as a second language. None of the listeners had any experience with languages with non-modally phonated vowels, except for English. Listeners who were either speakers of Spanish dialects with breathiness, or were exposed to those dialects (by family members who were native speakers) were excluded from the study. None of the listeners were familiar with the purpose of the study. All of the listeners were students at the University of California, Los Angeles, and all received $10 for their participation.

3.2.2 Stimuli

3.2.2.1 Languages and/or dialects

Although synthesized speech has proven to be useful in perception tasks, especially because it allows for more control over the stimuli, it is not clear what parameters should be synthesized at this stage in the research on the perception of phonation. In place of synthesized stimuli, naturally produced stimuli from 10 languages and/or dialects (Chong, Fuzhou, Green Hmong, White Hmong, Mon, Santa Ana del Valle Zapotec, San Lucas Quiaviní Zapotec, Tlacolula de Matamoros Zapotec, Tamang, and !Xóõ) were used. In the following section, I briefly discuss the languages/dialects used in this study, with emphasis on their phonation contrasts, though for some of these languages/dialects very little is known.

3.2.2.1.1 Chong

Chong is a Mon-Khmer language spoken in Thailand and Cambodia. Chong has four tones that are distinguished by phonation. Tone 1 is a level tone with modal phonation. Tone 2 is similar to tone 1, but has a higher pitch and ends with creak. Tone 3, a falling tone, is produced with breathy phonation. Tone 4 is also a falling tone, but is higher in pitch than tone 3 and ends with creak.

Chong breathy vowels are distinguished from modal ones by greater aspiration noise throughout the vowel, a larger H1-H2 value, and a decrease in the higher frequency energies (Blankenship 1997).

The recording used in this study was made by Theraphan Thongkum in 1986. The recording is of four male and four female speakers, between 50-60 years of age, of the Krathing dialect of Chong. The eight speakers produced each word, including some minimal sets, once in isolation.

3.2.2.1.2 Fuzhou

Fuzhou is a dialect of Min Dong, the form of Chinese spoken in the Fujian province of China. Fuzhou has seven tones (high level (55), high-falling (53), mid-level (33), low falling-rising (212), mid rising-falling (242), mid-rising (24), and high- level closed syllable (5)). The low falling-rising tone (212) is produced with breathy phonation and the mid-rising tone (24) is produced with slight creak (Lin, p.c.).

The recording used in this study was made by the author in 2004. The recording is of one 27-year-old male speaker of Fuzhou producing a minimal set in isolation.

3.2.2.1.3 Hmong

Though there are several dialects of Hmong, a Sino-Tibetan language spoken in Southeast Asia, the differences between the dialects are small and usually do not involve the breathy low tone. Thus, I will present a description of Hmong in general.

Hmong contrasts 7 tones. The low tone is produced with a breathy voice. The difference between H1 and H2, and the proportion of the glottal flow cycle taken up by the closed phase are successful measures for distinguishing phonations (Huffman 1985).

The recordings used in this study were made by the author in 2004. The recordings are of a 20-year-old male and a 21-year-old female speaker of Green Hmong and a 20-year-old male and 19-year-old female speaker of White Hmong. Each speaker produced a minimal set in isolation.

3.2.2.1.4 Mon

Mon is a Mon-Khmer language spoken in Thailand. There are two registers in Mon – a high-level tone with modal voice, and a low-level tone with breathy voice. However, Lee (1983) found that the difference between the registers was not primarily in phonation, but rather duration and pitch, with the second register being longer and lower than the first. Thongkum (1987) found evidence to the contrary. She concluded that a phonation difference does exist in Mon. Most of the speakers in her study produced the first register with a modal voice, and the second with a breathy voice. However, some speakers did not make this same distinction. For example, one speaker made a distinction between two degrees of breathiness, with the first register being less breathy than the second (Thongkum 1987).

The recording used in this study was made by Theraphan Thongkum in 1986. The recording is of several male and female speakers, between 50-60 years of age, of the dialect spoken in the Nakum Chum Sub-district, Rajaburi province, Thailand. The speakers produced each word, including some minimal sets, once in isolation.

3.2.2.1.5 Santa Ana del Valle Zapotec

Santa Ana del Valle Zapotec (SADV Zapotec) is an Otomanguean language spoken in Santa Ana del Valle, Oaxaca, Mexico. SADV Zapotec contrasts breathy, modal, and creaky vowels each associated with a tone. Breathy vowels have a slightly falling tone, modal vowels can be either high, or high-rising (without a change in phonation) and creaky vowels have a large falling tone. As previously mentioned, there is a gender difference in the production of phonation.

The recordings used in this study were made by the author over multiple sessions in 2002-2004. The recordings are of two males and one female between 40-60 years of age. The speakers produced each word, including minimal sets, in sentence-medial position, where phonation contrasts are most apparent for this language (Esposito 2003; 2005).

3.2.2.1.6 San Lucas Quiaviní Zapotec

San Lucas Quiaviní Zapotec (SLQ Zapotec) is an Otomanguen language spoken in San Lucas Quianviní, Oaxaca, Mexico. SLQ Zapotec contrasts modal, breathy, checked (a vowel followed by a glottal stop), and creaky phonation. It is also a tonal language, but the tones are dependent on the number and type of phonations found in a word. A characterization of the tonal system of SLQ Zapotec is as follows: modal vowels have a high tone, creaky vowels have a low tone, and phrase-final breathy vowels have an extra-low tone. The tone of checked and non-phrase-final breathy vowels is derived from the environment in which they occur (Munro and Lopez et al. 1999).

The recording used in this study was made by Matthew Gordon in 1999. The recording consists of one male and one female speaker producing a minimal set in isolation.

3.2.2.1.7 Tlacolula de Matamoros Zapotec

Tlacolula de Matamoros Zapotec (Tlacolula Zapotec) is an Otomanguen language spoken in Tlacolula de Matamoros, Oaxaca, Mexico. While there has not been much work on the phonation of Tlacolula Zapotec, it is believed to be very similar to SLQ

Zapotec (Lillehaugen, p.c.). Tlacolula Zapotec contrasts modal, breathy, checked, and creaky phonation. It is also a tonal language, but the tones are dependent on the number and type of phonations found in a word (Lillehaugen, p.c.).

The recording used in this study was made by Brook Lillehaugen in 2004. The recording consists of one male and one female speaker, producing a variety of sentences.

3.2.2.1.8 Tamang

Tamang is a Sino-Tibetan language spoken in Nepal. It contrasts four tones: high-falling, mid-high level, mid-low level, and very low. The two low-pitched tones have breathy phonation (Mazaudon 2003).

The recording used in this study was made by an unknown UCLA undergraduate in 1983. The recording consists of one male speaker producing words, including minimal sets, in isolation.

3.2.2.1.9 !Xóõ

!Xóõ is a Khoisan language spoken in Botswana and Namibia. It contrasts four tones (high, mid, low, and mid-falling) and five phonations (modal, breathy, creaky, strident, and pharyngealized) The phonations and tones are fully cross-classified. In addition, combinations of phonations can occur.

Breathy vowels are distinguished from modal ones by open quotient (as indicated by H1-H2), spectral tilt, and/or noise, depending on the speaker (Ladefoged et al. 1985).

The recording used in this study was made by Peter Ladefoged and Anthony Traill in the 1980's. The recording consists of several male and female speakers producing words, including some minimal sets, in isolation.

3.2.2.2 Manipulation of stimuli

The stimuli were composed of breathy vowels excised from real words consisting of an coronal stop + vowel. (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. To control for gender, only male voices were used in the experiment; female voices were used in the practice session.

Vowel quality was controlled in that only [a] was used. The vowels were cut right after the stop burst, leaving the CV formant transition, including any aspiration, 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 copy and pasting (or cutting, when natural vowel duration exceeded 250 ms) individual pulses. The original proportion of breathy to modal phonation was always preserved. For example, suppose the original vowel was 100 ms long, with 20 ms of breathy voice and 80 ms of modal voice. Vowel duration would be increased by cutting and pasting the individual pulses to create a vowel that was 250 ms, with 50 ms of breathy voice and 200 ms of modal voice.

F0 was normalized to a slightly falling f0 of 115-110 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. This is accomplished by first dividing the signal into smaller, separate, but overlapping, units. Next, these smaller units are altered by either repeating speech segments (when the desired pitch is higher than the original token) or omitting speech segments (when the desired pitch is lower than the original token). This step changes the duration and the f0 of the signal. In the last step, remaining smaller segments are overlapped and added together to recombine the speech signal. The resulting speech signal has the same spectrum as the original but with a different f0 (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 resynthesized only within a range of +/- 40 Hz from the original value. (For example, if the original f0 was 100 Hz, it would be resynthesized to a value in the range of 60-140 Hz.) This is because pilot data showed that resynthesizing within this range creates a < 1 dB change in the harmonics (Esposito 2006). (Expert listeners from the UCLA Phonetics Lab could not hear the < 1 dB difference between the natural tokens and the re-synthesized ones (Esposito 2006).)

As previously mentioned, spectral measures are sensitive to changes in formant structure. However, this was not problematic in the present case because the stimuli were always [a].

3.2.3 Phonetic measurements of the stimuli

Acoustic measures were made on all of the stimuli to determine which dimension(s) might correlate with listeners' perception. Only acoustic measures were made because the stimuli were only available as audio signals. CPP, H1-H2, H1-A1, H1-A2, H1-A3, the average of H1-H2 compared to A1, A2-A3, and H2-H4 were measured for all of the stimuli.

3.2.3.1 Periodicity

Periodicity was measured by taking the average height of the cepstral peak over the entire duration of the vowel. (In this case, the measure was made on the stimuli before the duration was manipulated.) Cepstral peak prominence was taken automatically using software based on Hillenbrand (1994) and created by the Bureau of Glottal Affairs, Division of Head/Neck Surgery, UCLA School of Medicine. This software is available for download at http://www.surgery.medsch.ucla.edu/glottalaffairs/index.htm.

3.2.3.3 Spectral measures

Because spectral measures reflect different aspects of phonation, it was necessary to take a variety of them. The following were taken: H1-H2, H1-A3, H1-A1, H1-A2, the average of H1-H2 compared to A1, A2-A3, and H2-H4. Measurements were made over a 30 ms window at one of two time points (either 50 ms after the start of the vowel or 50 ms before the end of the vowel) depending on the location of the non-modal phonation. Measurements were taken manually from an FFT. Spectrograms were used to position the 30 ms window.

3.2.4 Results of measures of stimuli

In section 3.2.4.1, we will examine a larger set of stimuli to help us characterize the phonations of each language/dialect as a whole. Then, in section 3.2.4.2, we will narrow the discussion to the phonations of the 40 stimuli used in the free-sort task. In the sections below, and throughout, I will discuss the hypothesized physical implications of the spectral measures based on the works of Holmberg et al. (1995) and Stevens (1977), though, in most cases, there has not been any work on the vibratory pattern of the vocal folds during phonation in the languages/dialects used in this experiment.

3.2.4.1 Overall phonation of languages/dialects used in this study

The results of the acoustic measures of phonation are presented in Figures 4-11. Each figure shows the average value for all the tokens by all the speakers. An asterisk marks the measures that were successful for each language. A measure was considered successful if there was a significant difference between the phonations in any direction. Significant differences were based on two-tailed *t*-tests. (It did not matter if the measure distinguished phonation in the expected direction. Any significant difference in phonation, even in the wrong direction, would still be a good cue for the free-sort task because, as will we see in section 3.3, listeners just have to distinguish the phonation types, not provide labels for them.)

Figure 4 is a graph of the CPP results. The mean height of the cepstral peak is plotted along the y-axis in dB. Languages and/or dialects are given on the x-axis. The lower the cepstral peak, the more breathy the stimuli. The arrow is pointing in the direction of increased breathiness.

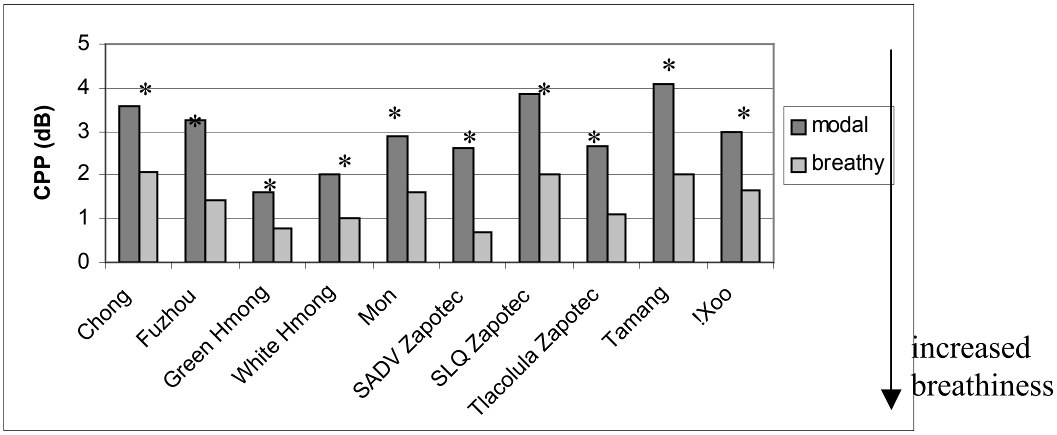


Figure 4: Graph of the average CPP value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

The results of the spectral measures are presented in Figures 5-11. The difference between the amplitudes of the harmonics (H1-H2, H1-A3, H1-A1, H1-A2, the average of H1-H2 compared to A1, A2-A3, and H2-H4) is given on the y-axis in dB. Languages and/or dialects are given on the x-axis. The greater the difference between the amplitude of the harmonics, the more breathy the stimuli. In each figure, an arrow is pointing in the direction of increased breathiness.

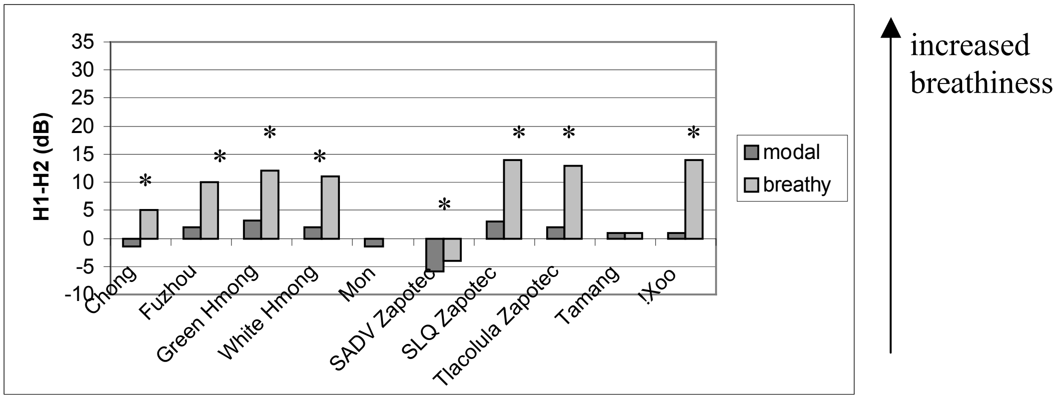


Figure 5: Graph of the average H1-H2 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

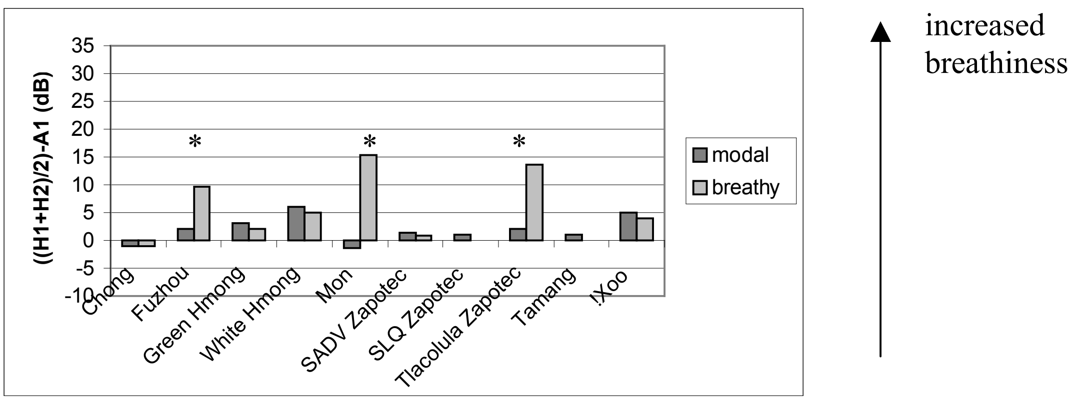


Figure 6: Graph of the average ((H1+H2)/2) – A1 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

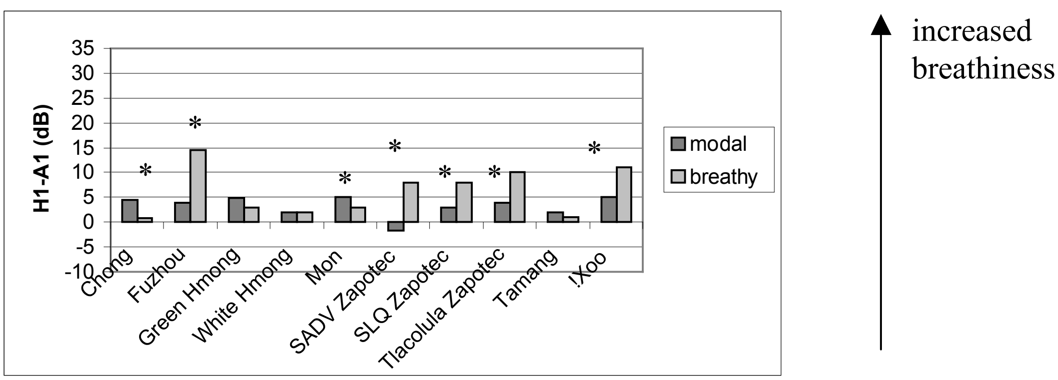


Figure 7: Graph of the average H1-A1 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

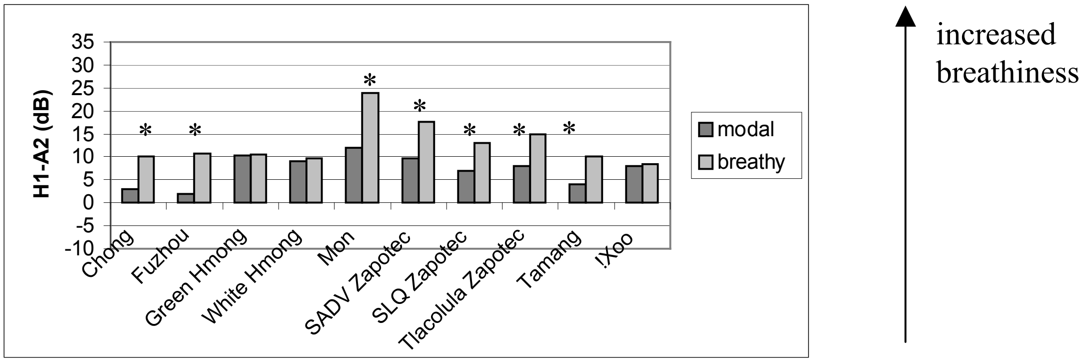


Figure 8: Graph of the average H1-A2 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

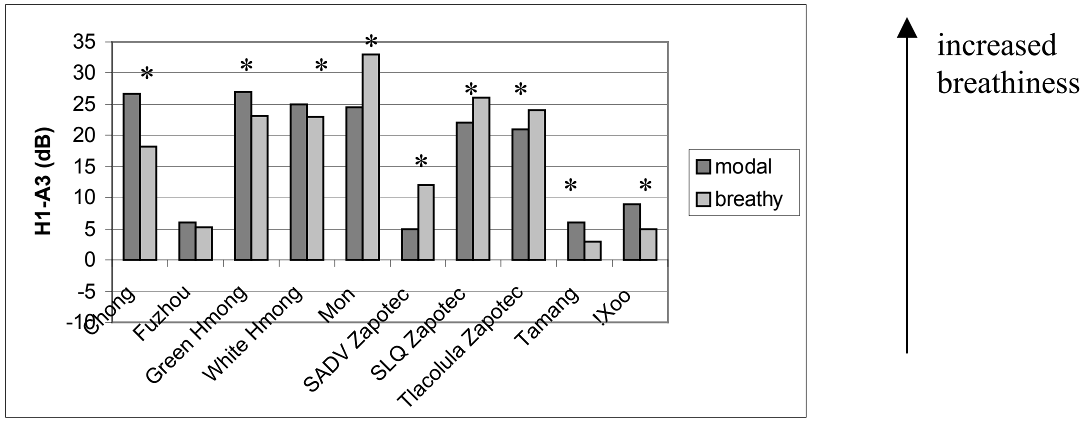


Figure 9: Graph of the average H1-A3 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

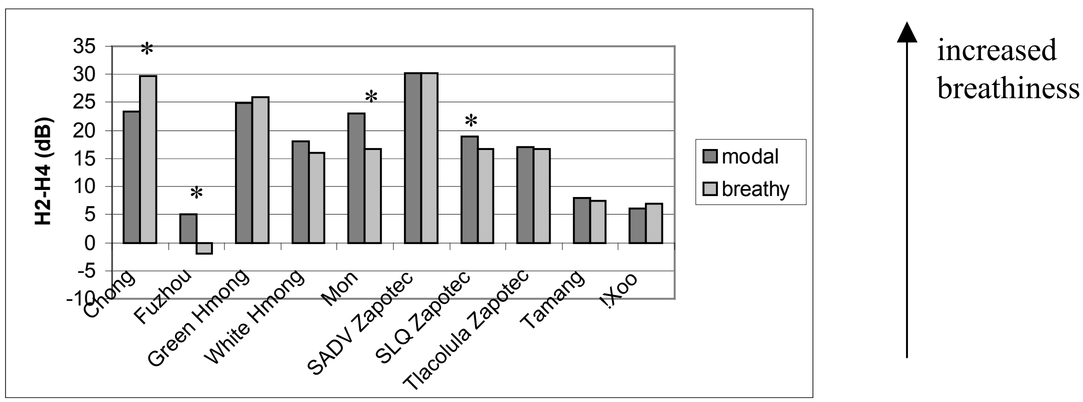


Figure 10: Graph of the average H2-H4 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

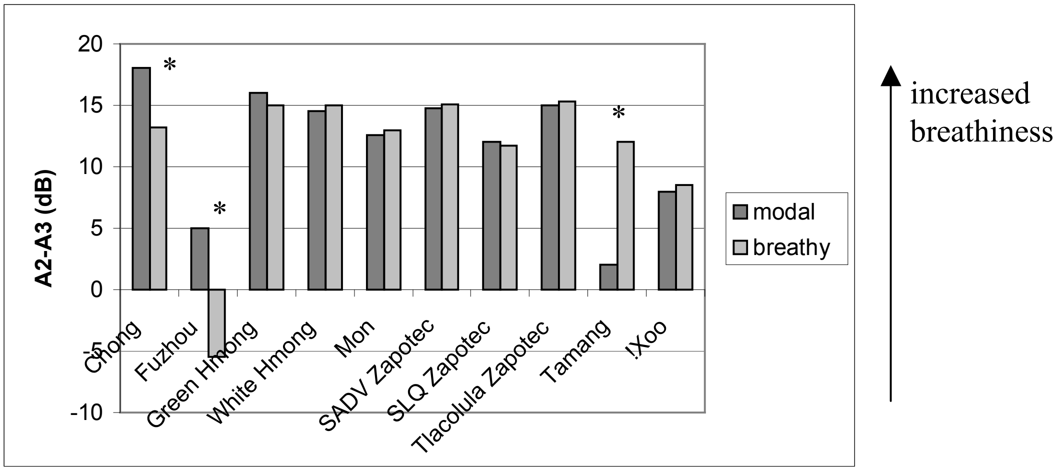


Figure 11: Graph of the average A2-A3 value for modal and breathy vowels in Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

Table 1 summarizes the information from the graphs presented in Figures 4-11. A check mark indicates which measures were successful for each language; a dashed line indicates that the measure was unsuccessful.

Languages and/or dialects	Measures							
	CPP	H1-H2	((H1+H2)/2) -A1	H1-A1	H1-A2	H1-A3	H2-H4	A2-A3
Chong	✓	✓	--	✓	✓	✓	✓	✓
Fuzhou	✓	✓	✓	✓	✓	--	✓	✓
Green Hmong	✓	✓	--	--	--	✓	--	--
White Hmong	✓	✓	--	--	--	✓	--	--
Mon	✓	--	✓	--	✓	✓	✓	--
SADV Zapotec	✓	✓	--	✓	✓	✓	--	--
SLQ Zapotec	✓	✓	--	✓	✓	✓	✓	--
Tlacolula Zapotec	✓	✓	✓	✓	✓	✓	--	--
Tamang	✓	--	--	--	✓	✓	--	✓
!Xóõ	✓	✓	--	✓	--	✓	--	--

Table 1: Success of the measures CPP, H1-H2, ((H1+H2)/2) -A1, H1-A1, H1-A2, H1-A3, H2-H4, A2-A3 for Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

From the table presented above, we can see that CPP is a good overall measure of phonation in all of the languages/dialects, suggesting that these ten languages/dialects all use a difference in noise to differentiate modal from breathy phonation. The second most successful measure is H1-A3, which works in all of the languages/dialects but Fuzhou, suggesting that all of the languages/dialects (except Fuzhou) use differences in the speed of vocal fold closure to distinguish phonation. The third best measure is H1-H2, which works in all of the languages/dialects except Mon and Tamang. (Note that H1-H2 *is* a successful measure of phonation in SADV Zapotec here, even though previous reports showed that this was not the case for male speakers (Esposito 2003; 2005). The reason for this discrepancy is that H1-H2 is a good measure for distinguishing breathy and modal phonation in SADV Zapotec, but not for distinguishing breathy, modal, *and* creaky phonation. Here, we are only considering modal and breathy phonation.) This suggests that all the languages/dialects (except Mon and Tamang) are using open quotient differences (the portion of time the vocal folds are open during each glottal cycle) to distinguish modal and breathy phonation. The fourth most successful measure in H1-A2, followed by H1-A1, and then by H2-H4. A2-A3 and ((H1+H2)/2) -A1 are the least successful measures of phonation in this sample.

In addition, multiple acoustic dimensions separate modal from breathy phonation in all of these languages/dialects. For example, in Chong, Fuzhou, SADV Zapotec, SLQ Zapotec, and Tlacolula Zapotec, CPP, H1-H2, and either H1-A1, H1-A2, and/or H1-A3 are all successful measures of phonation. This suggests that these languages/dialects are using some combination of noise, open quotient, and speed of vocal fold closure to produce phonation contrasts. As with other linguistic contrasts, redundancy is the rule here, not the exception.

3.2.4.2 Stimuli used in free-sort task

As previously mentioned, two breathy and two modal vowels were chosen from each language/dialects, producing a total of 40 stimuli. The results of the acoustic measures of the 40 stimuli (two modal, two breathy) are presented in Figures 12-19. No statistical comparison was made because there are only two modal and two breathy vowels for each language/dialect. In each figure, the arrow is pointing in the direction of increased breathiness.

Figure 12 is a graph of CPP. The height of the cepstral peak is plotted along the y-axis in dB. Figures 13-19 are of the spectral measures. The difference between the amplitudes of the harmonics (H1-H2, H1-A3, H1-A1, H1-A2, the average of H1-H2 compared to A1, A2-A3, and H2-H4) is given on the y-axis in dB. In all figures, languages/dialects are given on the x-axis.

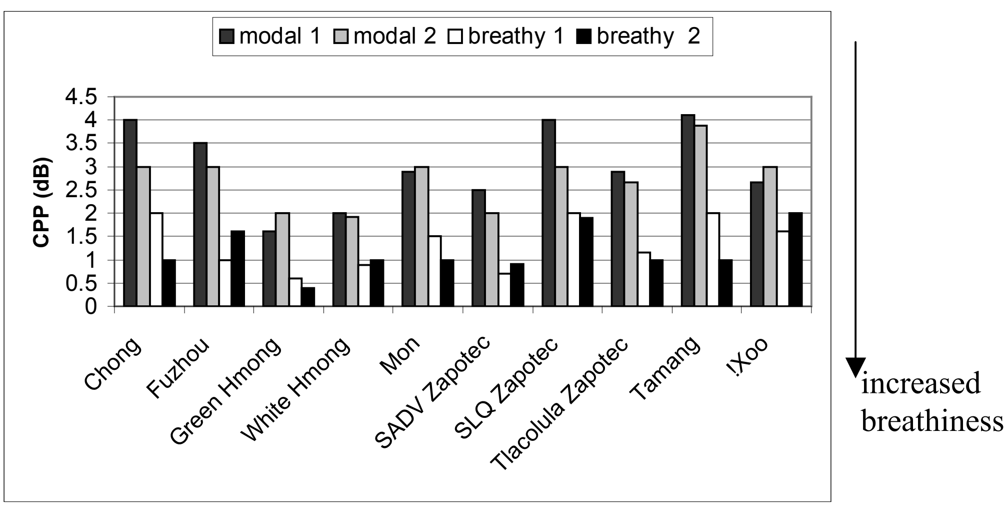
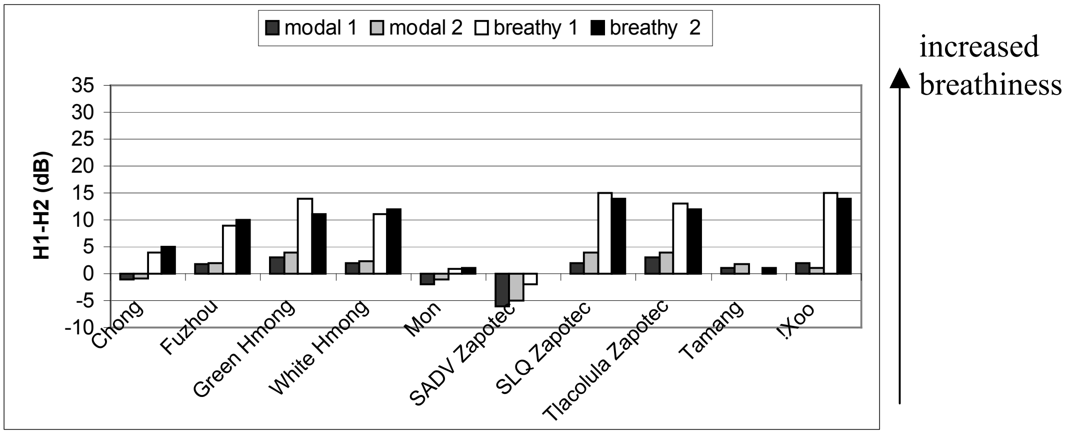


Figure 12: Graph of individual CPP values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).



.

Figure 13: Graph of individual H1-H2 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

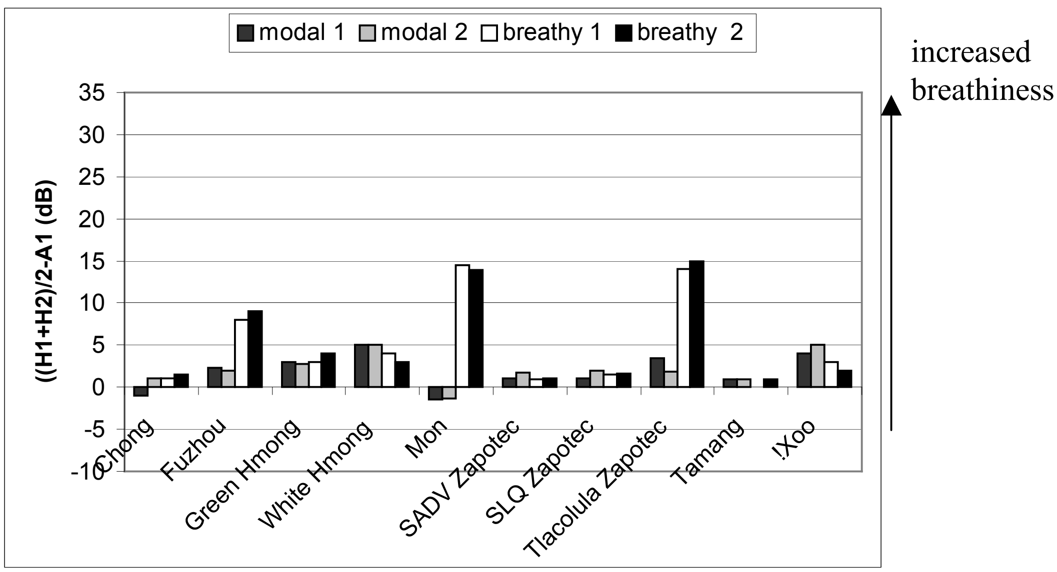


Figure 14: Graph of individual ((H1+H2)/2)-A1 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

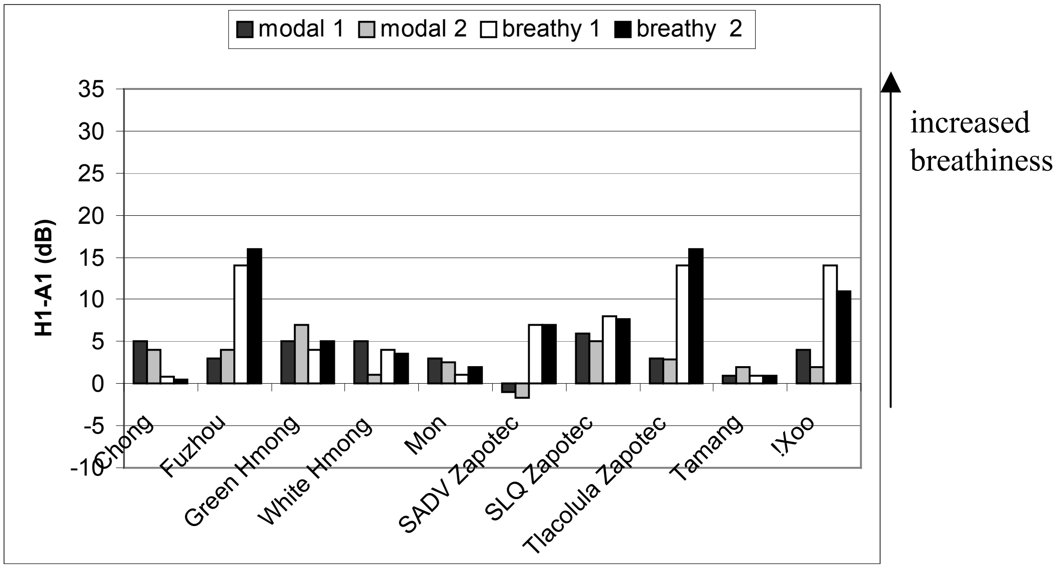


Figure 15: Graph of individual H1-A1 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

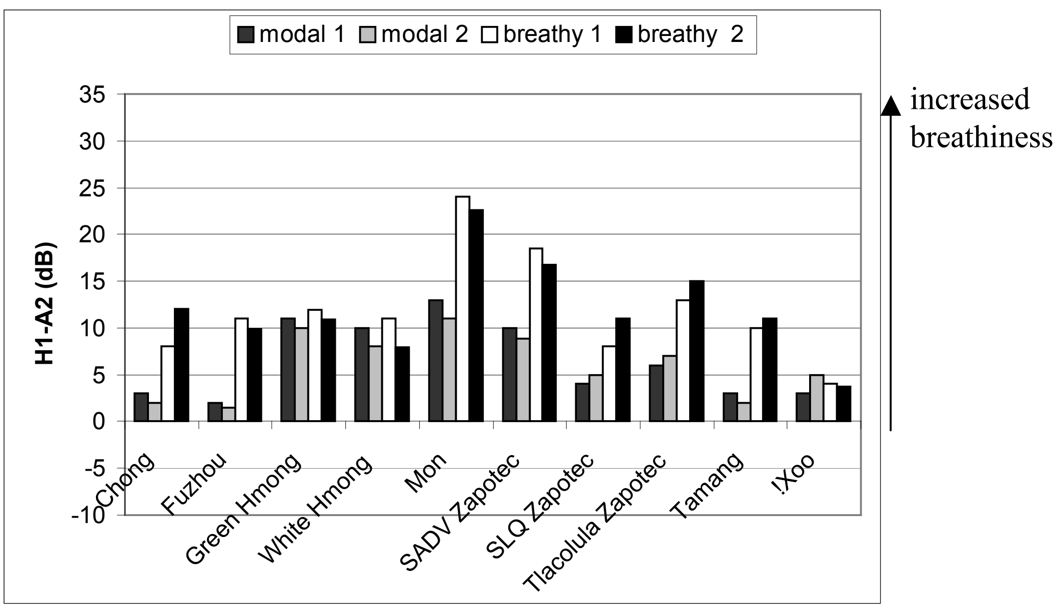


Figure 16: Graph of individual H1-A2 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

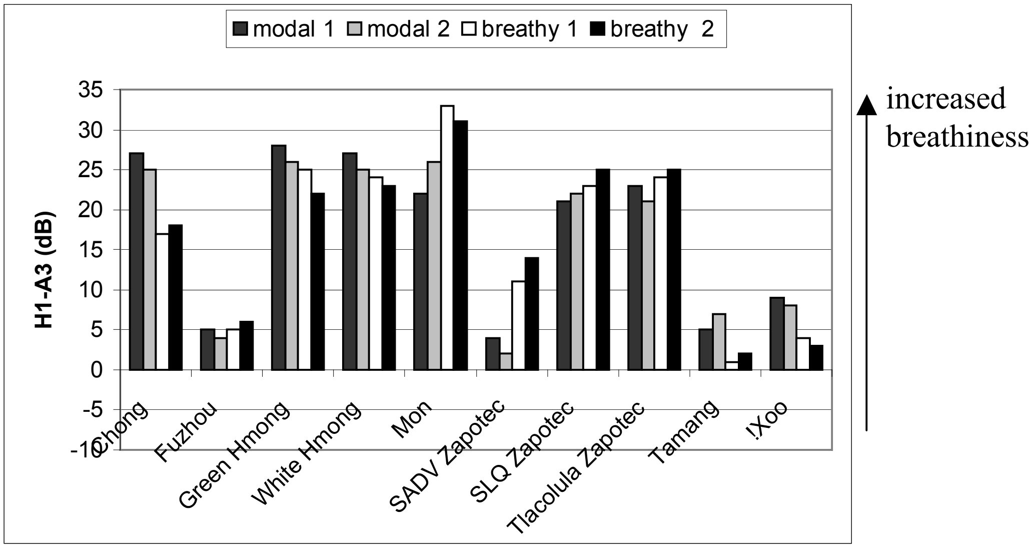


Figure 17: Graph of individual H1-A3 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

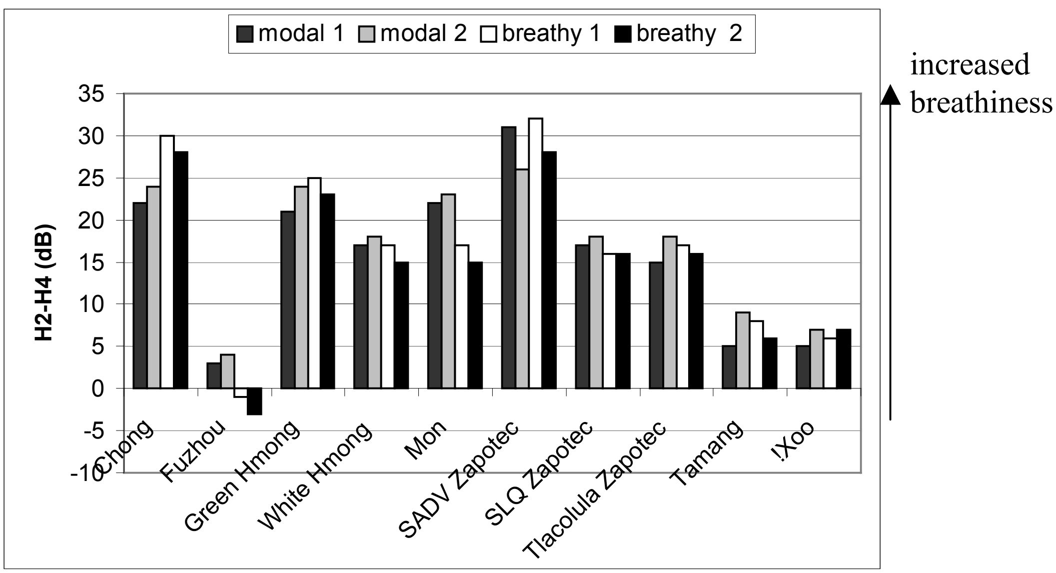


Figure 18: Graph of individual H2-H4 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

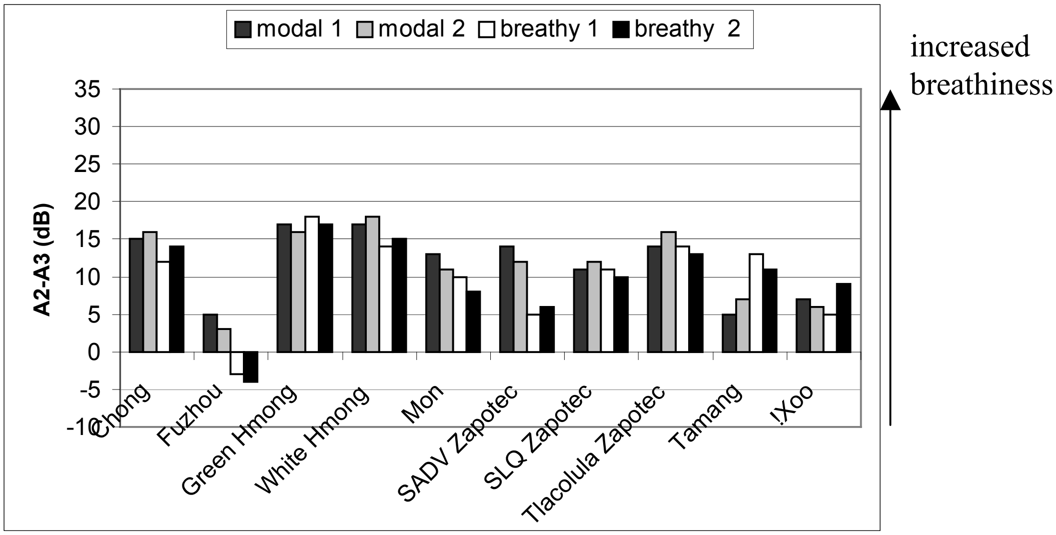


Figure 19: Graph of individual A2-A3 values for the 40 stimuli used in the free-sort task. There are four stimuli (two modal, two breathy) from each language (Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, Tamang and !Xóõ).

Table 2 summarizes the information from the graphs presented in Figures 12-19. A check mark indicates which measures successfully distinguish breathy and modal phonation for each language; a dashed line indicates that the measure was unsuccessful. Significance could not be determined by a *t*-test because four stimuli per language were not enough for a reliable *t*-test. Therefore, a measure was considered successful if both modal stimuli differed by at least 4 dB from both breathy stimuli on all measures except CPP. For CPP, a measure was considered successful if both modal stimuli differed by at least 1 dB from both breathy stimuli. (Pilot data showed that some 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).)

Languages and/or dialects	Measures							
	CPP	H1-H2	((H1+H2)/2) –A1	H1-A1	H1-A2	H1-A3	H2-H4	A2-A3
Chong	✓	✓	--	--	✓	✓	✓	--
Fuzhou	✓	✓	✓	✓	✓	--	✓	✓
Green Hmong	✓	✓	--	--	--	--	--	--
White Hmong	✓	✓	--	--	--	--	--	--
Mon	✓	--	✓	--	✓	✓	✓	--
SADV Zapotec	✓	✓	--	✓	--	✓	--	✓
SLQ Zapotec	✓	✓	--	--	--	✓	--	--
Tlacolula Zapotec	✓	✓	✓	--	✓	✓	--	--
Tamang	✓	--	--	--	✓	✓	--	✓
!Xóõ	✓	✓	--	✓	--	✓	--	--

Table 2: Success of the measures CPP, H1-H2, ((H1+H2)/2) -A1, H1-A1, H1-A2, H1-A3, H2-H4, A2-A3 for the 40 stimuli from Chong, Fuzhou, Green Hmong, White Hmong, Mon, SADV Zapotec, Tlacolula Zapotec, Tamang and !Xóõ.

For the most part, success of the various measures on the 40 stimuli was similar to each language as a whole (Table 1). However, there were some exceptions to this. For example, H1-A3 was successful for all of the languages except Fuzhou, but in the 40 individual stimuli, this measure was not successful for Green and White Hmong, in addition to Fuzhou. Furthermore, ((H1+H2)/2) -A1 and A2-A3 were successful for the stimuli from Mon and SADV Zapotec, respectively, but these measures were not successful in distinguishing phonations in these particular languages as a whole.

One important issue to address is whether or not these eight measures are sufficient to categorize phonation in these ten languages/dialects. Discriminant analysis, a procedure that determines which variables discriminate between two or more groups, was used to determine if these eight 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 that these eight acoustic measures accounted for 91% of the variance in the data, with four measures doing most of the work: CPP accounting for most of the variance (46% of the variance), followed by H1-H2 (27% of the variance), H1-A2 (10% of the variance), and then H1-A3 (8% of the variance). Therefore, it is likely that Spanish listeners will use CPP to make their judgments about these stimuli because it accounted for the most variance in the data.

3.3 Procedure

3.3.1 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 just heard, but English and Spanish listeners do not have distinct categories to identify. (In the literature on English /ɹ/ and /l/ perception by Japanese listeners, for example, studies generally rely on the fact that Japanese listeners have learned some English and, therefore, could identify English words.) One possible solution would be to train the English and Spanish speakers 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.

Another solution would be to use a discrimination task instead of an identification one. This way, all three listener populations could be asked the same question without training. However, a discrimination task is undesirable because discrimination involves less specifically linguistic processing than identification. Identification tasks involve matching the stimulus to a stored mental representation, and, thus, are more likely to reflect the influence of the listener's language background. For example, Ingram and Park (1998), in their experiment on the perception of Australian /ɹ/ and /l/ by Korean and Japanese speakers, found an effect of listeners' language background in an identification task, but not a discrimination one. 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 and Eng 2003). In the VSR task, subjects were first asked to sort stimuli, which were presented as icons on a computer screen, by moving an icon along a vertical dimension labeled "high" (for high-frequency) at one end, and "low" (for low-frequency) at the other. Subject had the ability to move the stimuli anywhere along this vertical dimension. After the sorting, subjects were asked to rate the stimuli along the vertical dimension (which was, essentially, a visual analog scale). Granqvist and Eng (2003) compared the results of the VSR task to that of the visual analog scale. Results showed that VSR was a better and more reliable procedure than the more traditional method. The VSR method can be adapted for categorization, as a two-way sort with no scale.

The free-sort or VSR method is advantageous because a reference to a category does not need to be made to sort the stimuli. Thus, all three listening populations could 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 and Pruitt 1995).)

3.3.2 Design of the free-sort task

The free-sort task used in this study consisted of stimuli presented as icons on a computer screen. The free-sort task was implemented in Matlab for Windows (the script (Wilson 2005) is available for download at www.linguistics.ucla.edu/faciliti/facilities/perception/matlab.html) and run on a laptop with Sennheiser HD 280 professional headphones and a Telex p500 external soundcard. All the stimuli were presented in one trial and arranged in a random order. One way the free-sort task used in this experiment differed from the VSR task previously mentioned is that, in this experiment, the screen contained two "sorting boxes". 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 2006). Figure 20 is a schematic representation of the free-sort task.

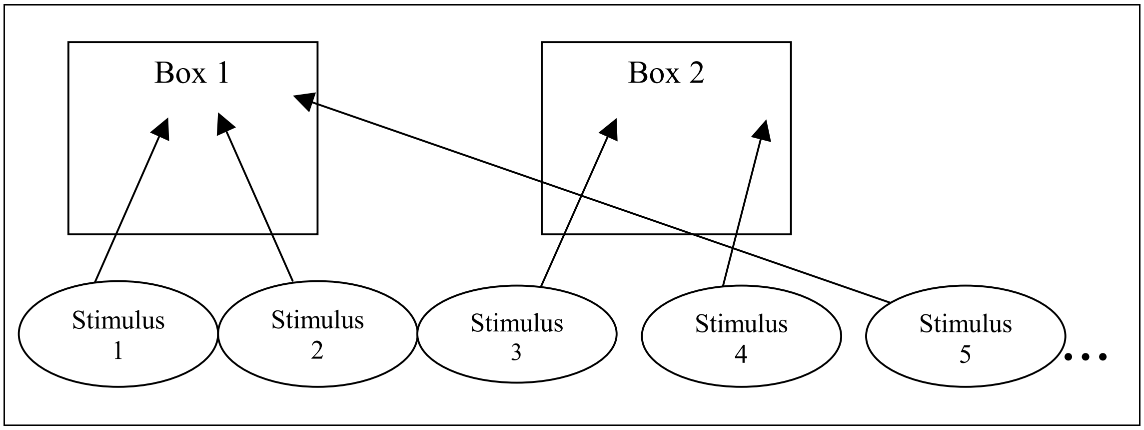


Figure 20: Schematic representation of the free-sort task. Arrows represent one possible sorting of the stimuli. The ellipsis represent additional stimulus.

Listeners began with a practice session composed of 20 stimuli produced by female talkers chosen from a subset of the languages/dialects used in the experiment. (The languages/dialects in the practice session were: Chong, White Hmong, Green Hmong, Mon, SADV Zapotec, SLQ Zapotec, Tlacolula Zapotec, and !Xóõ. Fuzhou and Tamang were not used in the practice because female voices were not available for these languages/dialects.) The directions for the experiment were given in the listeners' native languages to direct the listener to the phonation contrasts in his/her language. (Studies, such as Elman et al. 1977, have shown a shift in performance in identification tasks depending on the language that the instructions were given in.) For English and Spanish listeners, the directions were written and given orally in the subject's native language. For Gujarati, it was not possible to find a native speaker who was able to run the experiment. Therefore, for Gujarati listeners, the directions were given in English, during which time subjects could ask questions about the experiment. Then, the instructions were given orally in Gujarati. (The directions were pre-recorded in Gujarati and played back to listeners at the start of the experiment.) Because many Gujaratis living in the US are not literate in Gujarati, the directions were only given orally. In addition, Gujarati listeners were shown a picture and were asked to say three spontaneous sentences describing it aloud in Gujarati just before the start of the experiment. This was done to put these listeners into a Gujarati mindset and reduce English interference.

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 to not leave an item alone in a box. (This was to ensure that Spanish listeners, who might perceive all the stimuli to be the same, would not sort everything into one box.)

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 2006).

Listeners were told to play all the stimuli 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. On average, the entire free-sort task took 20 minutes.

3.4 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 was each pair put into the same sorting box?). All possible pairings of stimuli (462 pairs in all) were examined. Per-pair consistency, correlation between acoustic properties and listeners' judgments, and the percent of pairs that listeners correctly identified were calculated and will be discussed in sections 3.4.1, 3.4.2, and 3.4.3, respectively.

3.4.1 Per-pair consistency

Per-pair consistency was determined by looking at how listeners' responses deviated from randomness. (Per-pair consistency does not deal with the possibility that listeners put two random stimuli into one box, and the rest of the stimuli in the other box. In any case, however, individual responses were examined, and no listener behaved in this manner.) More specifically, per-pair consistency was determined by looking at the difference between those who said "same" (or the who said "different") vs. 50% (where 50% is taken to represent random responses across the group). (Here, "same" means that the stimuli were put in the same box and "different" means that the stimuli were put in different boxes.) Thus, the consistency value can lie between 0 (that is, no different from 50%) and 50 (that is, maximally different from 50%), and the larger the 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.)

For example, if there were 10 listeners from Language X and 3 stimuli (A, B, and C), the results might look like this:

Stimuli Pairs	"Same" responses	"Different" responses	Deviation from 50-50
stimulus A vs. B	5	5	0%
stimulus A vs. C	6	4	10%
stimulus B vs. C	4	6	10%

Table 3: An example of how to calculate per-pair consistency.

In this example, the average deviation from a 50-50 split is (0+10+10)/3 = 6.7%: Language X speakers are close to evenly split on all stimulus pairs.

3.4.1.1 Results

The average per-pair consistency by language is presented in the table below.

Listeners	Average Consistency
English	21.05%
Spanish	20.93%
Gujarati	48.8%

Table 4: Average per-pair consistency for English, Spanish, and Gujarati listeners.

The distribution range of the per-pair consistencies for English, Spanish, and Gujarati listeners is presented in Figures 21, 22, and 23, respectively. In all three figures, the x-axis represents the per-pair consistencies (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.

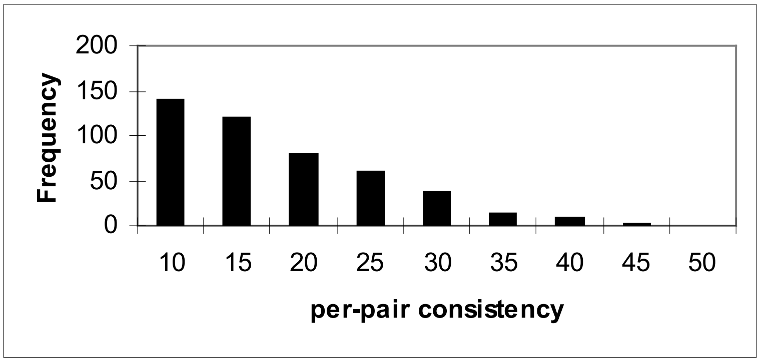


Figure 21: Distribution range of per-pair consistencies for English listeners.

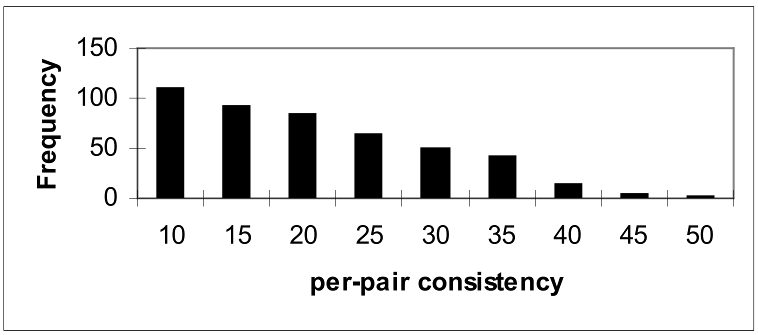


Figure 22: Distribution range of per-pair consistencies for Spanish listeners

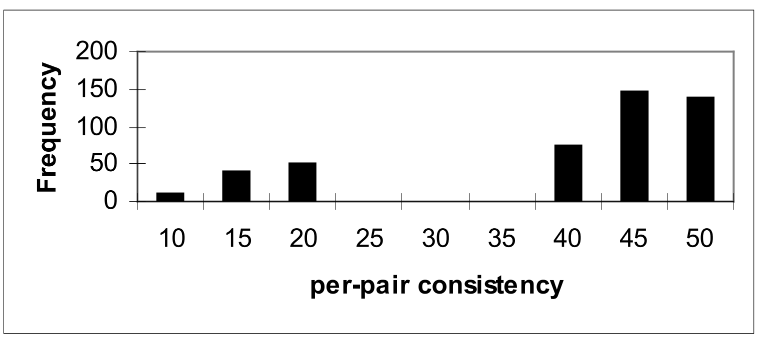


Figure 23: Distribution range of per-pair consistencies for Gujarati listeners.

Results showed that there was a lack of consistency within the English and Spanish listener groups. English listeners only deviated 21.05% from randomness, with per-pair consistencies in the range of approximately 10% to 45%, with most in the 10% range (Figure 21). Furthermore, Spanish listeners only deviated 20.93% from randomness, with per-pair consistencies in the range of approximately 10% to 50% with most in the 10% range (Figure 22). Gujarati listeners' judgments were more consistent than either English or Spanish listeners. Gujarati listeners deviated 48.8% (out of a total of 50%) from randomness, with per-pair consistencies in the range of 10% to 50%, with most $\geq$ 40% (Figure 23).

In the next section, we will see if the listeners' judgments correlated with any acoustic measures.

3.4.2 Correlations

Correlations between acoustic properties and listeners' judgments were determined for each of the eight acoustic measures again based on pairs of stimuli. Listeners judgments were again coded as putting each pair of stimuli into the same box or not. For each pair, their acoustic difference (in dB) was taken for each of the eight acoustic measure. 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 (referred to as "H1-H2 Diff." in the graphs below) between these two stimuli is 5 dB. All possible pairings of stimuli along each of the acoustic dimensions (CPP, H1-H2, H1-H2, H1-A3, H1-A1, H1-A2, the average of H1-H2 compared to A1, A2-A3, and H2-H4) were examined. 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.

3.4.2.1 Results for English listeners

The correlations between the measures and perceptual judgments for English listeners are presented in Figures 24-31. Eight different measures are presented in eight graphs per language. In these figures, the y-axis represents the perceptual judgments (that is, the number of listeners who agreed that this pair has the same phonation). The x-axis represents the difference (in dB) of the two stimuli in question. The r value and the slope of the line are given for each graph. A significant correlation between the acoustic measure and the perceptual judgment is marked with an asterisk in the upper-right hand corner of each graph.

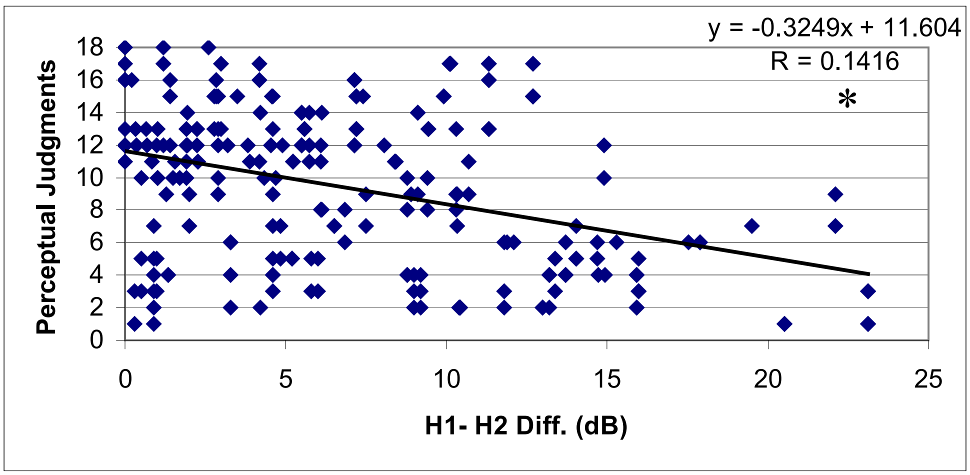


Figure 24: Correlation of the H1-H2 difference between the stimuli pairs and English listeners' judgments.

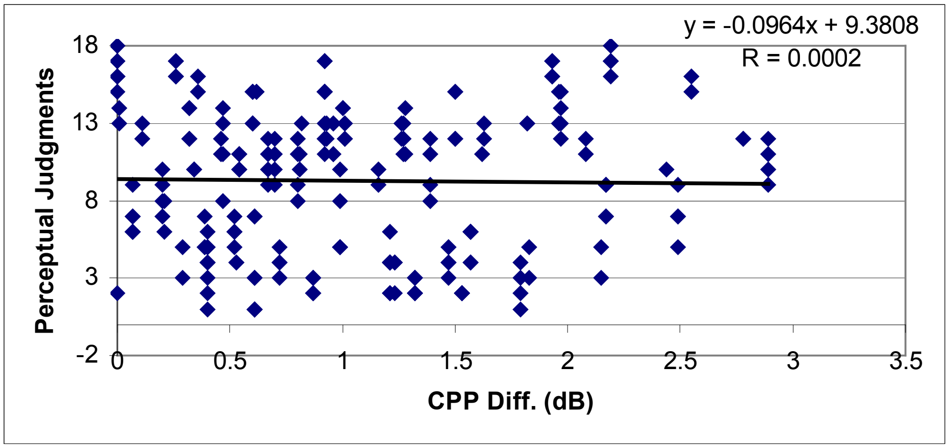


Figure 25: Correlation of the CPP difference between the stimuli pairs and English listeners' judgments.

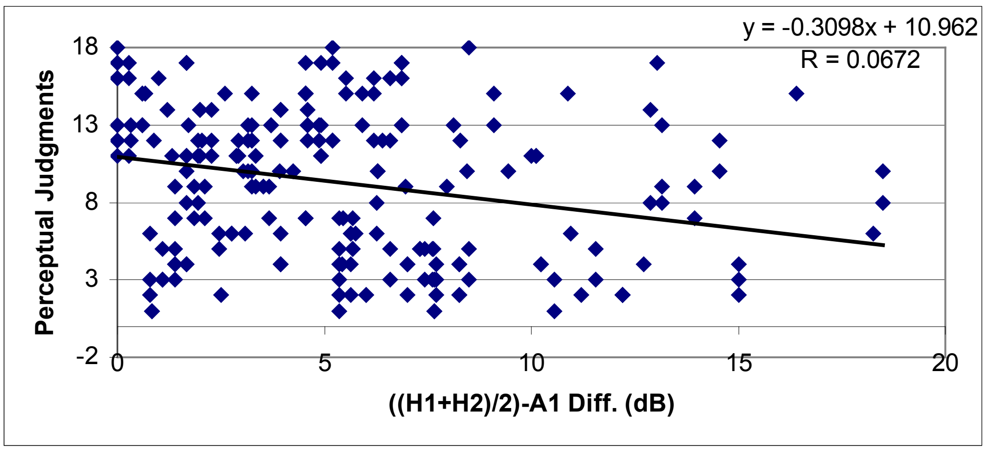


Figure 26: Correlation of the ((H1+H2)/2) -A1 difference between the stimuli pairs and English listeners' judgments.

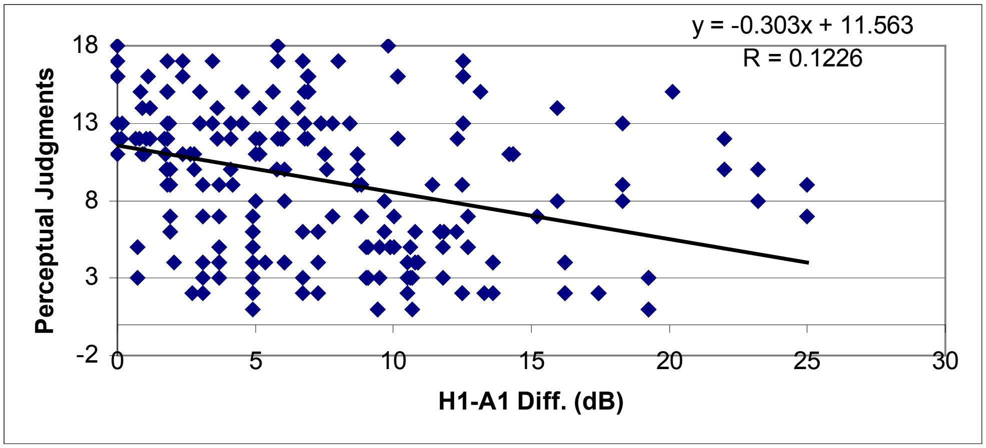


Figure 27: Correlation of the H1-A1 difference between the stimuli pairs and English listeners' judgments.

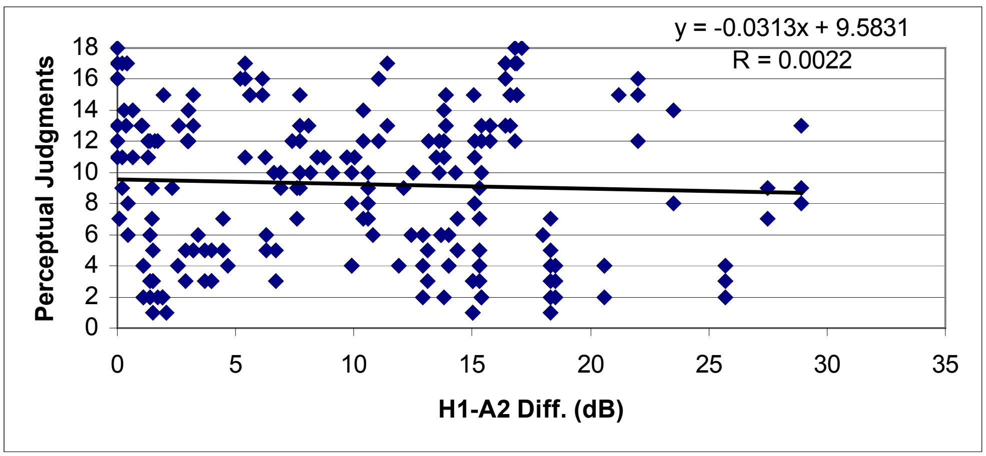


Figure 28: Correlation of the H1-A2 difference between the stimuli pairs and English listeners' judgments.

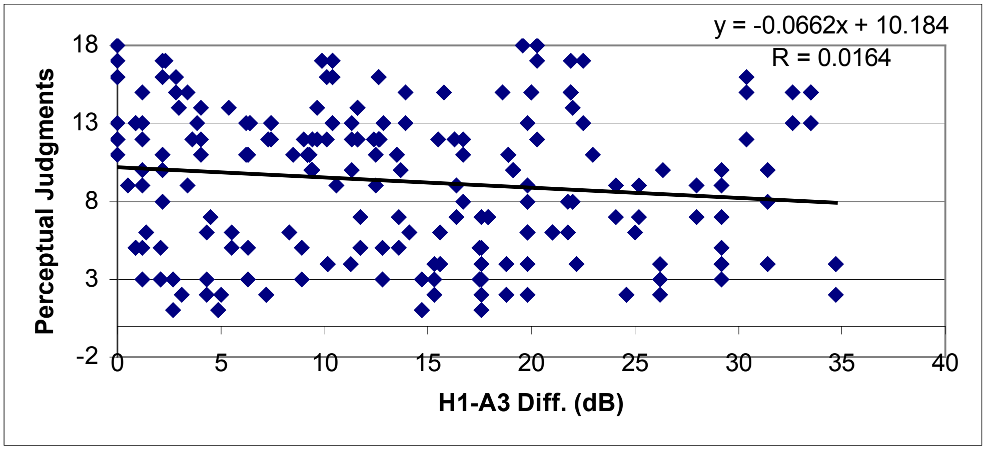


Figure 29: Correlation of the H1-A3 difference between the stimuli pairs and English listeners' judgments.

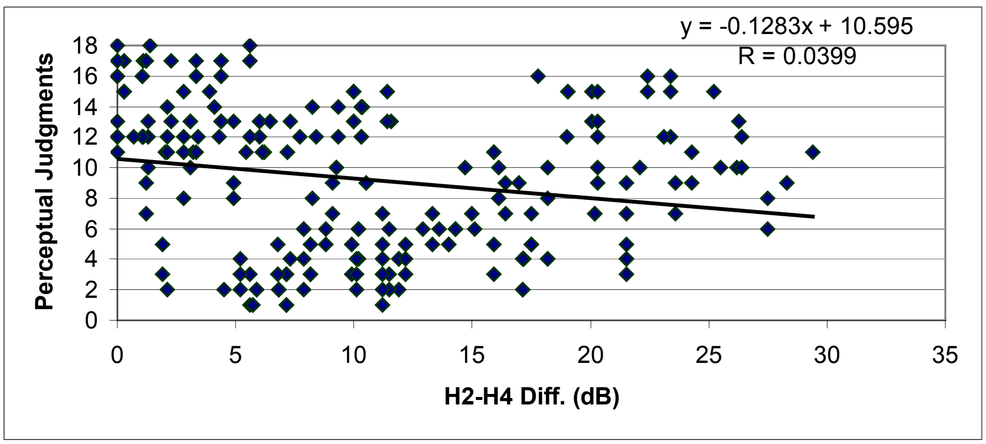


Figure 30: Correlation of the H2-H4 difference between the stimuli pairs and English listeners' judgments.

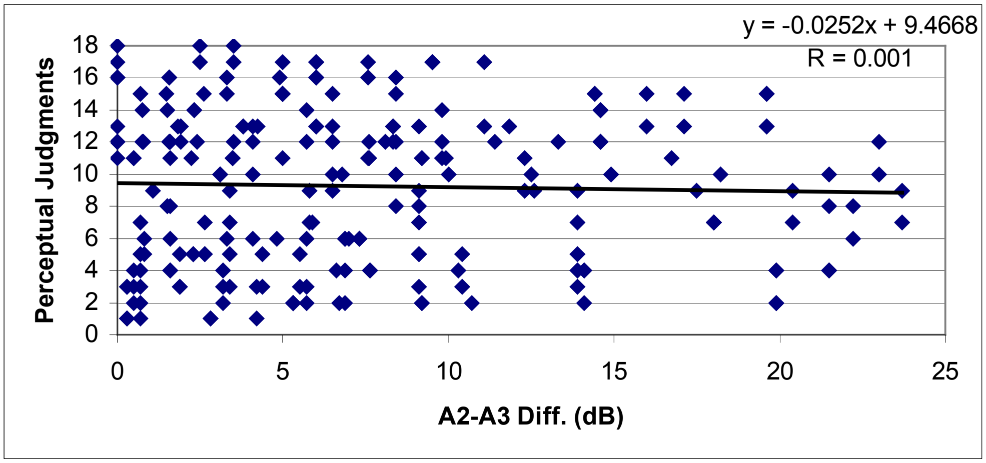


Figure 31: Correlation of the A2-A3 difference between the stimuli pairs and English listeners' judgments.

Overall, results for English listeners showed a significant but weak relationship between H1-H2 and the perceptual judgments. There was no significant relationship between the other measures and the English listeners' judgments.

3.4.2.2 Results for Spanish listeners

The correlations between the measures and perceptual judgments for Spanish listeners are presented in Figures 32-39. As with the figures presented above, the y-axis represents the perceptual judgments and the x-axis represents the difference (in dB) of the two stimuli in question. A significant correlation between the acoustic measure and the perceptual judgments is marked with an asterisk in the upper right-hand corner of each graph.

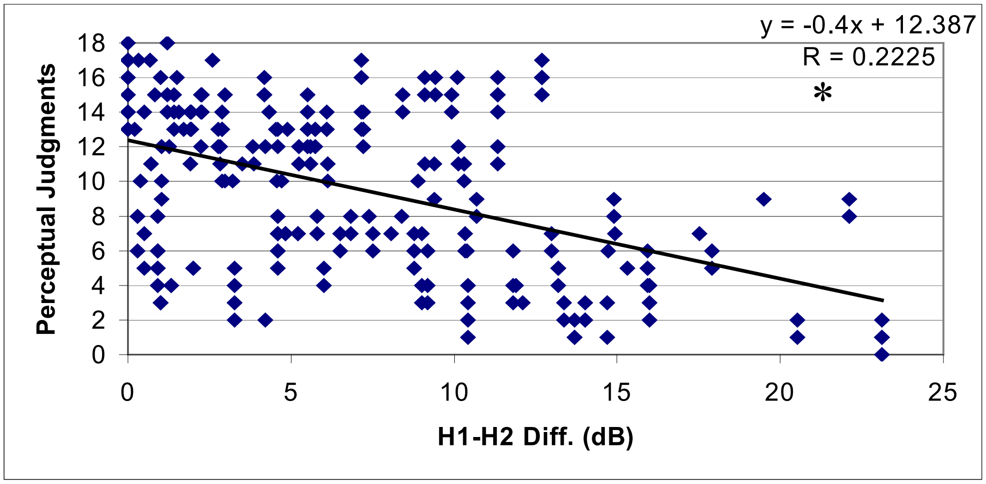


Figure 32: Correlation of the H1-H2 difference between the stimuli pairs and Spanish listeners' judgments.

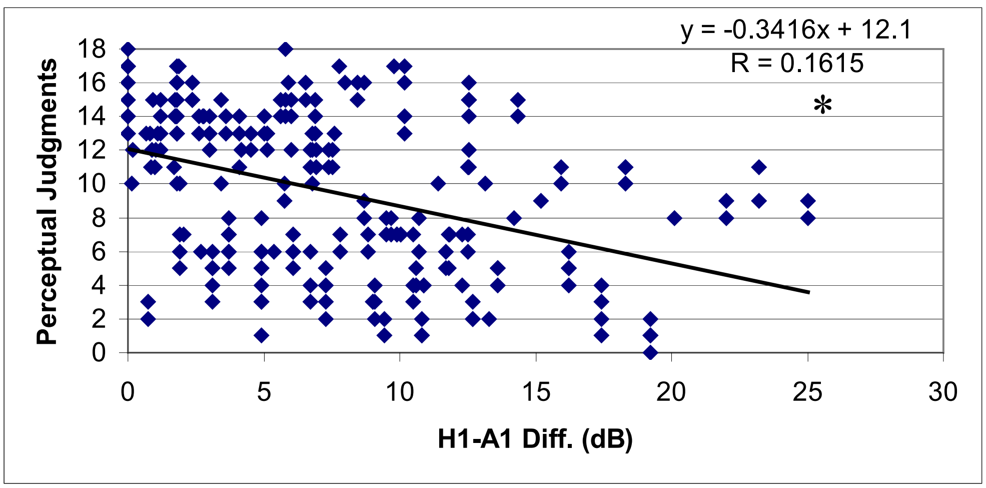


Figure 33: Correlation of the H1-A1 difference between the stimuli pairs and Spanish listeners' judgments.

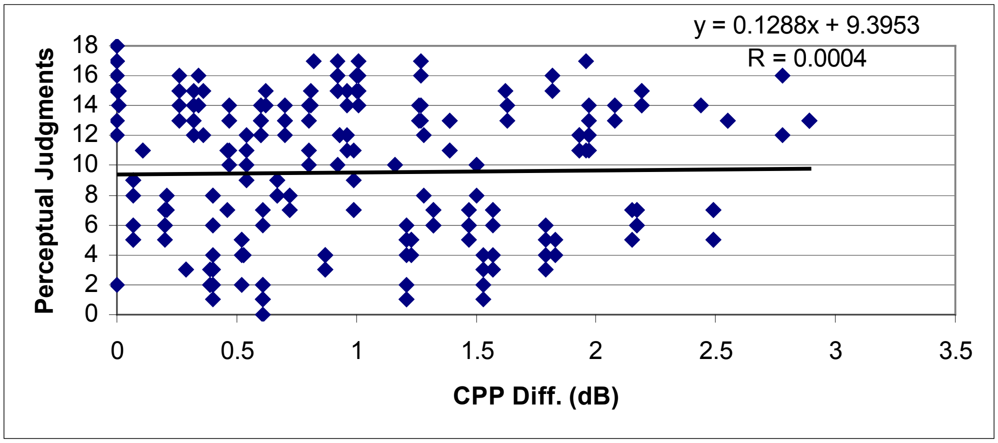


Figure 34: Correlation of the CPP difference between the stimuli pairs and Spanish listeners' judgments.

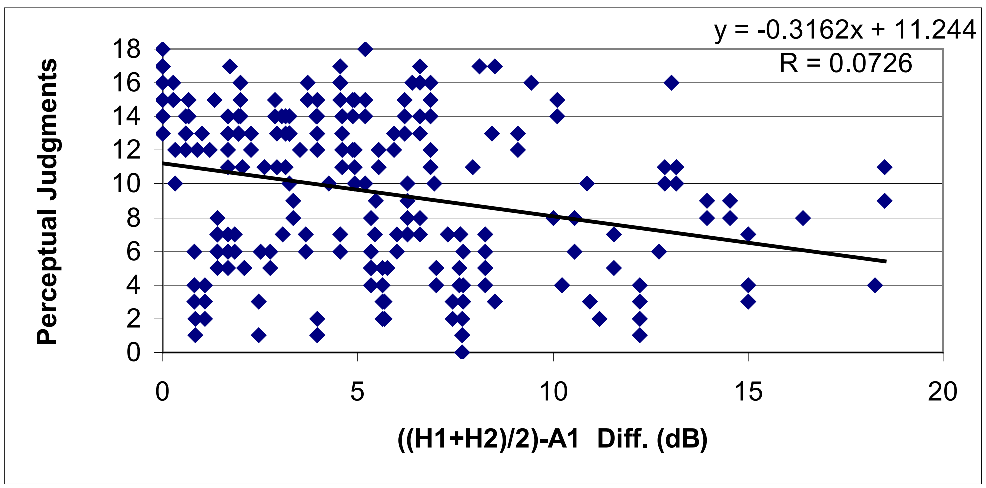


Figure 35: Correlation of the ((H1+H2)/2)-A1 difference between the stimuli pairs and Spanish listeners' judgments.

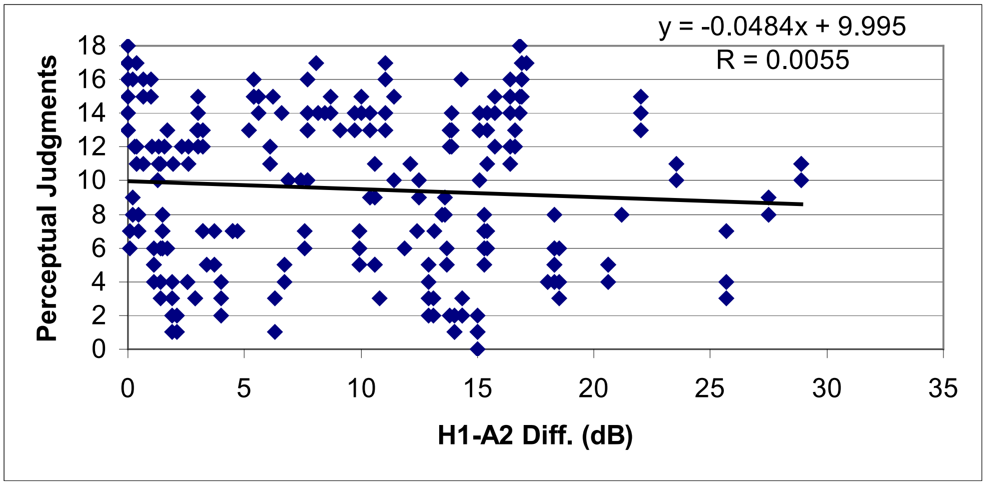


Figure 36: Correlation of the H1-A2 difference between the stimuli pairs and Spanish listeners' judgments.

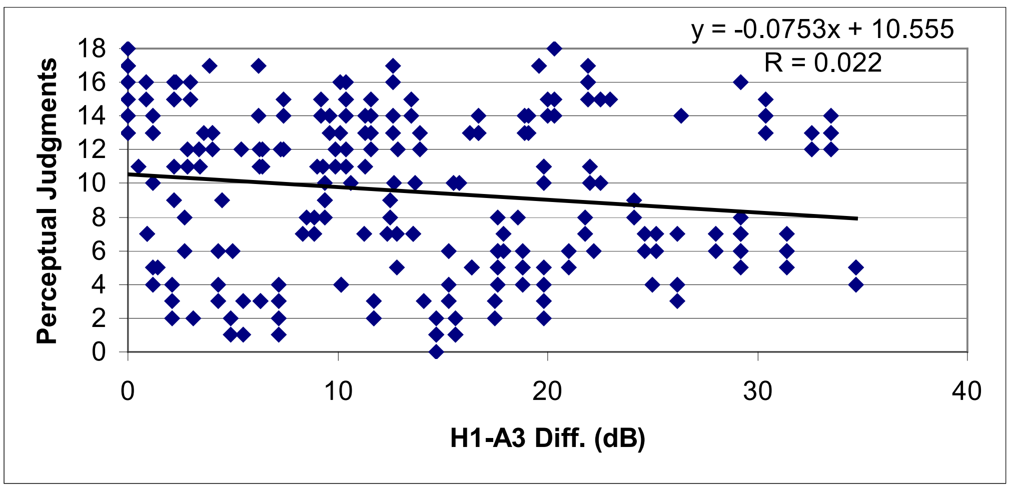


Figure 37: Correlation of the H1-A3 difference between the stimuli pairs and Spanish listeners' judgments.

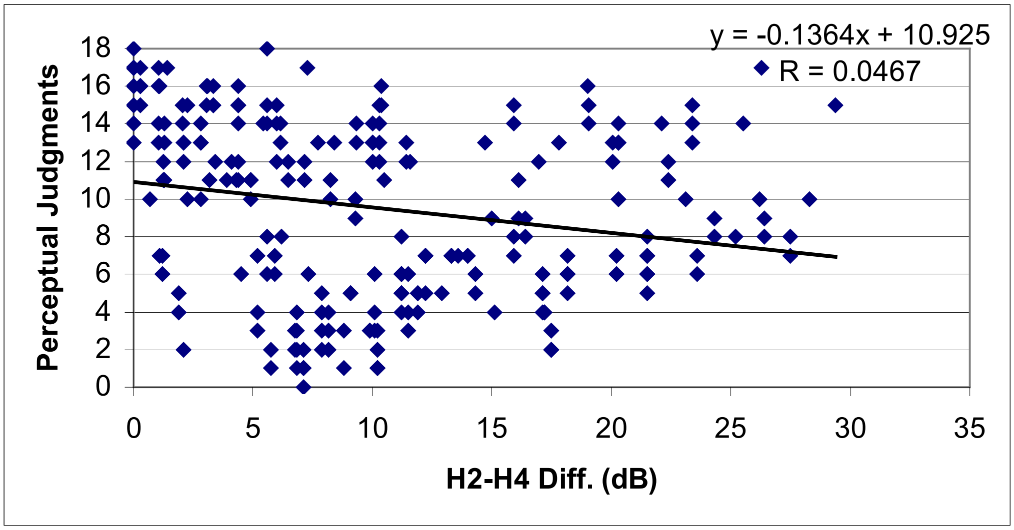


Figure 38: Correlation of the H2-H4 difference between the stimuli pairs and Spanish listeners' judgments.

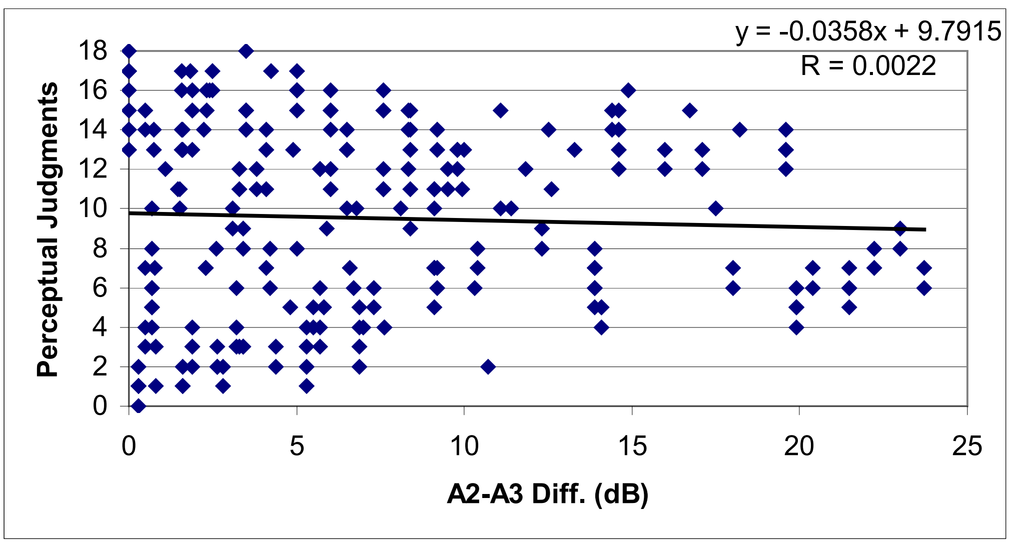


Figure 39: Correlation of the A2-A3 difference between the stimuli pairs and Spanish listeners' judgments.

In sum, results for Spanish listeners showed a significant but weak relationship between H1-H2 and also H1-A1 and Spanish listeners' judgments. There was not a significant relationship between any of the other measures and the Spanish listeners' judgments.

3.4.2.3 Results for Gujarati listeners

The correlations between the measures and perceptual judgments for Gujarati listeners are presented in Figures 40-47. Again, the y-axis represents the perceptual judgments and the x-axis represents the difference (in dB) of the two stimuli in question. A significant correlation between the acoustic measure and the perceptual judgments is marked with an asterisk in the bottom right-hand corner of each graph.

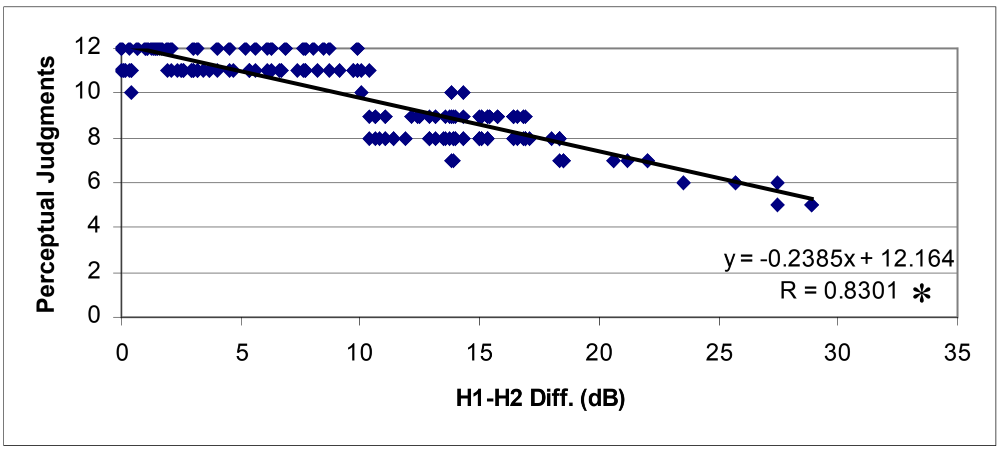


Figure 40: Correlation of the H1-H2 difference between the stimuli pairs and Gujarati listeners' judgments.

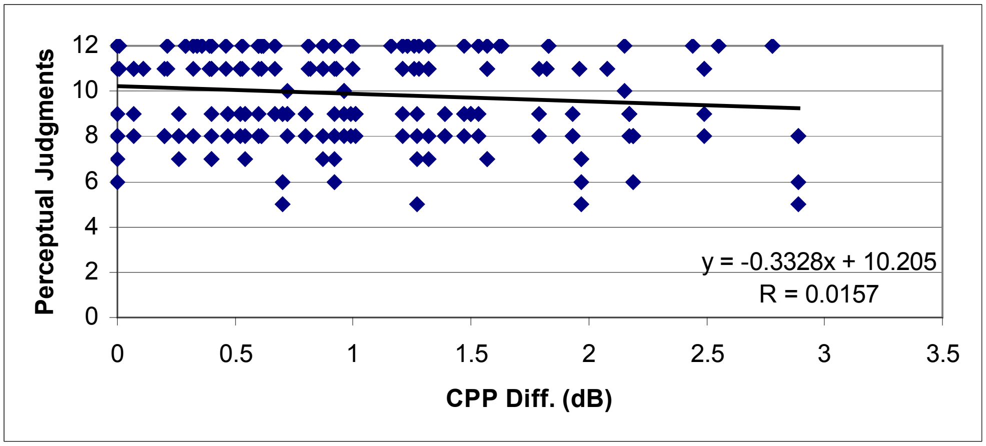


Figure 41: Correlation of the CPP difference between the stimuli pairs and Gujarati listeners' judgments.

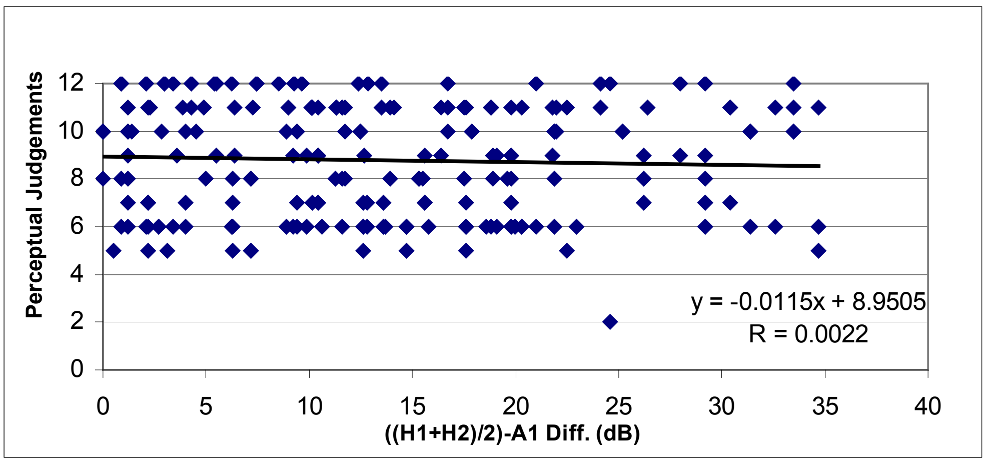


Figure 42: Correlation of the ((H1+H2)/2)-A1 difference between the stimuli pairs and Gujarati listeners' judgments.

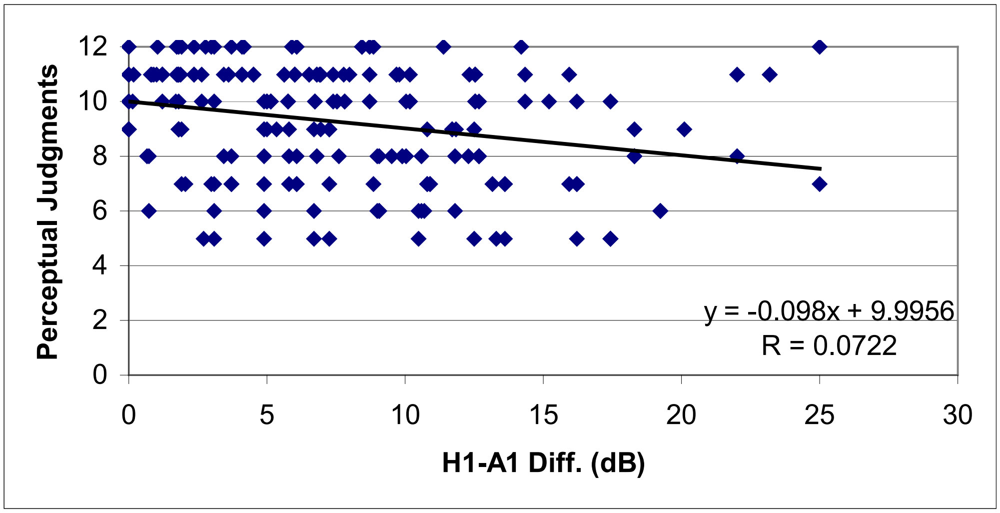


Figure 43: Correlation of the H1-A1 difference between the stimuli pairs and Gujarati listeners' judgments.

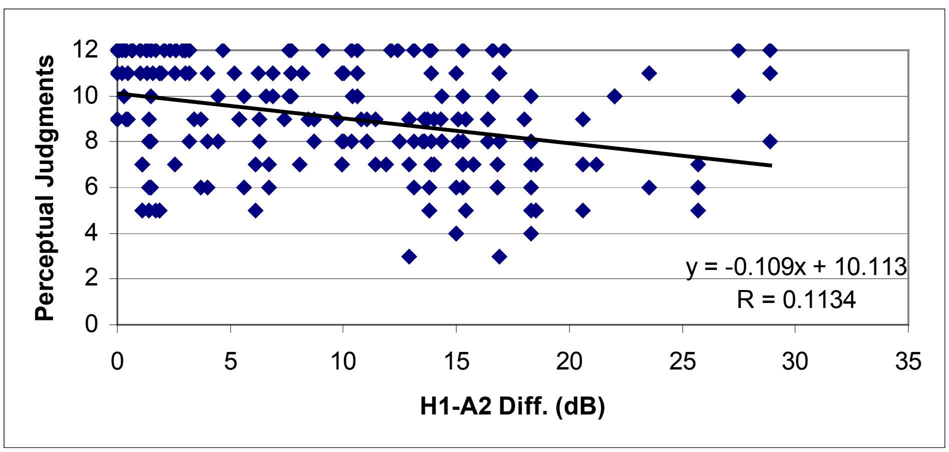


Figure 44: Correlation of the H1-A2 difference between the stimuli pairs and Gujarati listeners' judgments.

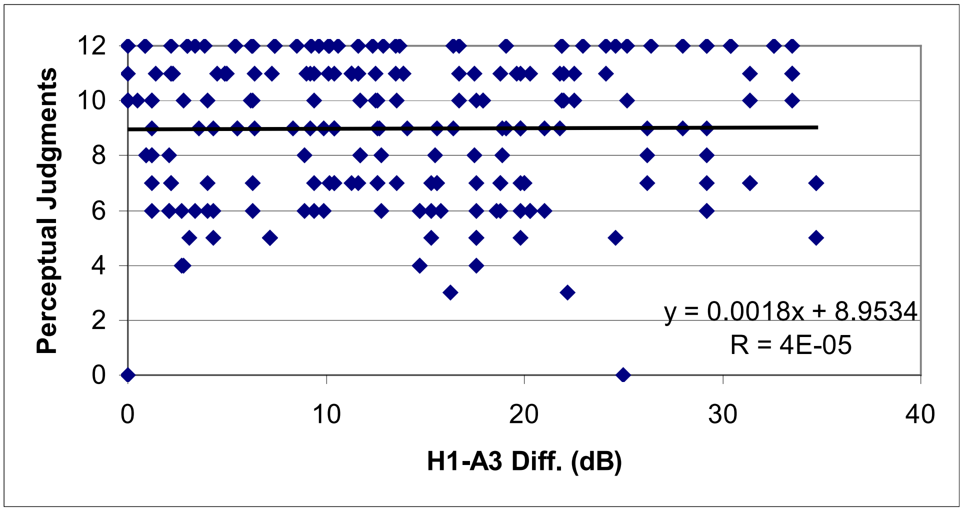


Figure 45: Correlation of the H1-A3 difference between the stimuli pairs and Gujarati listeners' judgments.

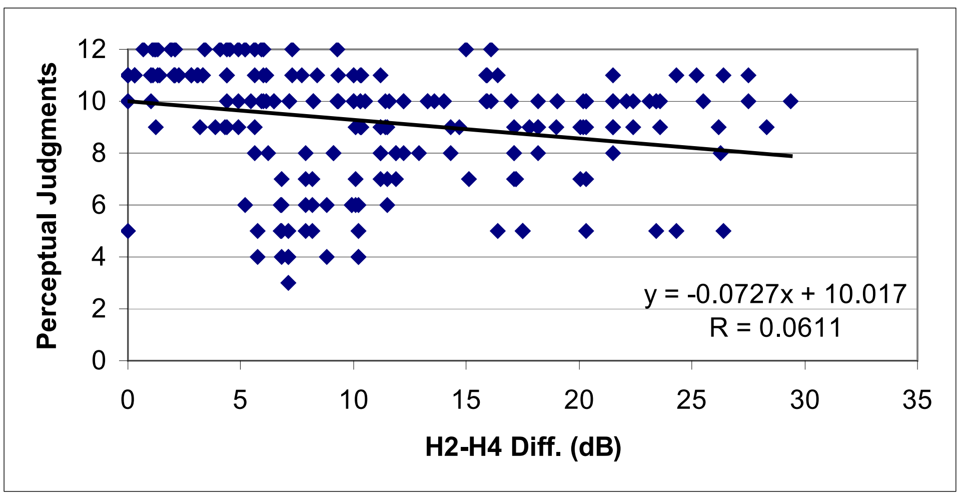


Figure 46: Correlation of the H2-H4 difference between the stimuli pairs and Gujarati listeners' judgments.

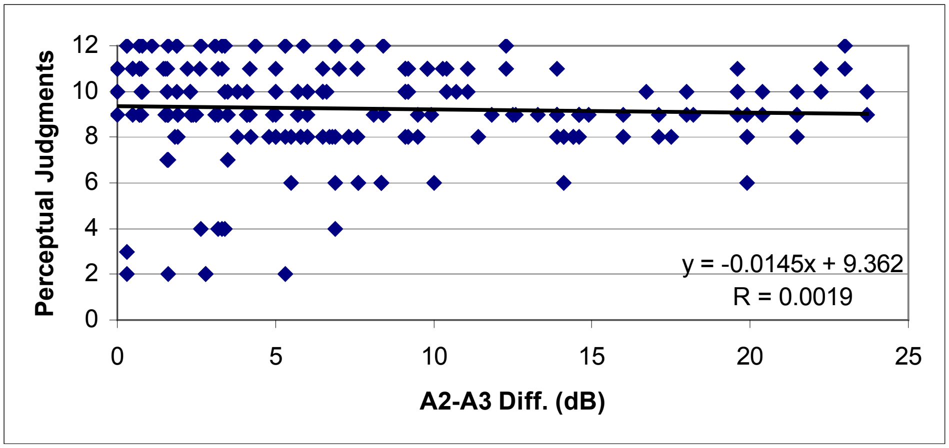


Figure 47: Correlation of the A2-A3 difference between the stimuli pairs and Gujarati listeners' judgments.

There was a strong, significant relationship between H1-H2 and Gujarati listeners' judgments. There was no significant relationship between any other measures and the Gujarati listeners' judgments.

3.4.3 Percent correct

Because the stimuli were taken from real language samples, it was easy to divide the stimuli into two phonation categories, breathy or modal, based on the phonemic categorization of the stimuli in their source languages and/or dialects. Using this natural division of the stimuli, it was possible to determine the percent of the time that listeners sorted the stimuli "correctly". This is interesting from the point of view that language experience should hurt a listener's ability to perceive a similar contrast in a different language if the cues are different. Thus, it will be interesting to see if the Gujarati listeners are able to correctly identify breathy and modal vowels that are not produced along H1-H2 differences (because the correlation analysis showed the Gujarati listeners only pay attention to H1-H2).

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:

Percent correct:

$$\frac{\text{\# of times listeners paired stimuli with the same phonation category in the same box}}{\text{Total number of pairs}} \times 100$$

All possible pairings of the stimuli were examined.

3.4.3.1 Results

The percent correct for each listener population is presented in the table below.

Listeners	Percent Correct
English	54%
Spanish	58%
Gujarati	90%

Table 5: Percent correct for English, Spanish, and Gujarati listeners.

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.

3.5 Summary

The experiment in this chapter centered on the following questions:

1) How does linguistic experience affect one's perception of phonation?
2) What acoustic properties correlate with listeners' perception?

It was hypothesized that: 1) Gujarati listeners' judgments would be more uniform than either English or Spanish listeners, 2) that they would be more successful in distinguishing modal from breathy vowels, and 3) that they would base their judgments on H1-H2. The results of this experiment confirmed these hypotheses. Gujarati listeners were more consistent and did better at sorting the stimuli than either English or Spanish listeners. In addition, Gujarati listeners' judgments were strongly correlated with H1-H2.

It was predicted that Spanish listeners would perceive phonation differences poorly and lack consistency in their judgments. It was also predicted that the Spanish listeners would rely on the largest auditory difference between the stimuli, in this case CPP, which accounted for most of the variance in the stimuli. Results showed that Spanish listeners were not very consistent in their judgments, nor could they correctly sort modal from breathy phonation. However, there was a pattern to their judgments. Listeners' sorting of the stimuli was significantly correlated with H1-H2 and H1-A1. It is not clear why Spanish listeners used H1-A1 and H1-H2 and not CPP, which the discriminant analysis showed was the best measure to distinguish phonation in this stimulus set. This issue will be revisited in the next chapter.

Finally, for English listeners, studies on the perception of allophonic variation suggested that English listeners would be sensitive to breathy phonation while studies on the perception of pathologically-disordered phonation suggested that English listeners would distinguish phonations poorly with lack of agreement. In addition, it was predicted that the English listeners would base their judgments on H1-H2 and CPP. Results showed that English listeners were not very consistent in their judgments, nor could they correctly distinguish modal from breathy phonation. Despite the allophonic breathiness in English, listeners' judgments were neither more consistent nor more accurate than speakers of a language with only modal phonation (Spanish). In fact, English and Spanish listeners' results were remarkably similar. However, there was a pattern to the English listeners' judgments. As predicted, listeners' sorting of the stimuli was significantly correlated with H1-H2, but they did not use CPP.

Even though CPP was the most successful measure in distinguishing breathy from modal phonation (of the eight measures tested), there was not a significant relationship between this measure and judgments for any of the three listener groups. A possible explanation comes from looking at the CPP values of the stimuli (Figure 4) the range of values is very narrow (from .5- 4 dB). Perhaps, listeners are not sensitive to such small differences along this stimulus dimension, but additional research will need to be conducted before this can be confirmed.

Chapter 4: Similarity-rating task

4.1 Introduction

In the last chapter, we saw that linguistic experience does affect one's perception of phonation in that Gujarati listeners did behave more consistently than English and Spanish listeners. We also saw that, in terms of acoustic measures, there was a significant correlation between H1-H2 and English listeners' judgments, H1-H2 and H1-A1 and Spanish listeners' judgments, and H1-H2 and Gujarati listeners' judgments. However, while there was a significant correlation between H1-H2 and judgments for all three listener groups, only the Gujarati listeners' judgments were *strongly* correlated with H1-H2.

In this chapter, we again explore the questions:

1) How does linguistic experience affect listeners' perception of phonation?
2) What acoustic properties correlate with listeners' perception?

However, this time, listeners were asked to participate in three similarity-rating tasks, where they had to explicitly judge the similarity of pairs of Mazatec vowels. Results were analyzed using multidimensional scaling (MDS), which is ideal for exploring a perceptual space when the number and kinds of dimensions that underlie the stimuli are not well known beforehand, as is the case with phonation (see Johnson 2003 for a more detailed description of multidimensional scaling).

Multidimensional scaling has been used in numerous speech perception experiments. A few examples of phonetic properties that have been explored using MDS are: voice quality (Kempster et al. 1991; Kreiman et al. 1990; Kreiman et al. 1992;

Kreiman et al. 1993; Kreiman and Gerratt 1996, etc.), tone (Gandour and Harshman 1978), vowel quality (Terbeek 1977, Rakerd and Verbrugge 1985, Fox et al. 1995), fricatives (Weiner and Singh 1974), and /ɹ/ and /l/ (Iverson and Kuhl 1996).

One of the goals of this chapter is to see whether listeners classify the stimuli into categories and how they rate them for similarity. In addition, analyzing the results using MDS will give us a more detailed picture of the acoustic cues that listeners are using to make their judgments. (In the free-sort task, listeners could only sort the stimuli into one of two boxes, while in the similarity-rating tasks, they can make finer judgments on the stimuli.) An additional goal is to try to replicate the results of the previous experiment with a different set of stimuli. The question arises, will the listeners use the same acoustic cues for rating similarity as they used for classification (the free-sort task) even with a different stimuli set? It is hypothesized that once again all three listener groups will use H1-H2 to make their judgments. In addition, it is expected that Spanish listeners will also use H1-A1, as they did in the free-sort task.

The organization of this chapter is as follows: section 4.2 will explain the methods used in this study, including a description of the listeners (section 4.2.1) and the stimuli (section 4.2.2.), section 4.3 will explain the general procedure for the similarity-rating task, section 4.4 will be a presentation of the results, and section 4.5 will be a summary of the chapter.

4.2 Methods

4.2.1 Listeners

The same listeners from the experiment in Chapter 3 participated in the three similarity-rating tasks.

4.2.2 Stimuli

4.2.2.1 Jalapa Mazatec

In this experiment, listeners were judging the similarity of pairs of stimuli. Because pair-wise comparisons were made, it was desirable to control for as many variables as possible. The most straightforward way to do this was to use stimuli produced by a single speaker of a language with contrastive phonation in each similarity-rating task. Naturally produced stimuli were drawn from Jalapa Mazatec (hereafter Mazatec), an Otomanguean language spoke in San Felipe Jalapa Diaz, Oaxaca, Mexico. Mazatec was chosen for a number of reasons: it contrasts breathy, modal, and creaky phonation (as well as low, mid, and high tone, but tones and phonations are fully cross-classified), it has been the subject of a number of studies on phonation (e.g., Kirk et al. 1993, etc.; Blankenship 1997), and extensive recordings were available from the UCLA Phonetics Archive.

Mazatec breathy vowels are characterized by an initial portion of breathiness followed by a more modal phonation for the remainder of the vowel. Breathiness is maintained for about 33% of the vowel (about 50 ms) (Blankenship 1997). In addition, breathy vowels are less periodic and have a lower cepstral peak than modal vowels. Mazatec breathy and modal are best distinguished by the measures H1-H2, H1-A2, and cepstral peak prominence (CPP). This success of these three measures suggests that Mazatec breathy vowels are produced by differences in the open quotient (which correlates with H1-H2), abruptness of vocal fold closure (which correlates with H1-A2) and aperiodicity (which correlates with CPP).

In addition, there are gender differences in the production of phonation. Blankenship (1997) showed that female speakers produce breathy vowels with significantly larger H1-H2 value than males do at approximately 50-125 ms into the vowel. However, there was no difference on the measures H1-A2 and CPP.

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.

4.2.2.2 Manipulation of the Stimuli

The stimuli were composed of breathy vowels excised from real words consisting of a coronal stop + vowel. (As with the previous experiment, coronals were chosen because there was more data available for this place of articulation.) The vowels with the best quality audio recording were chosen from three of the 12 Mazatec talkers, producing a total of 48 stimuli (eight breathy and eight modal stimuli from each of the three talkers). Of the three talkers, one was male (Talker 1); the other two were female (Talkers 2 and 3).

As with the first experiment, vowel quality was controlled in that only [a] was used. Vowels were cut right after the stop burst, leaving the stop transition intact.

Each vowel was normalized to the average duration (250 ms.) per talker using the same procedure explained in section 3.2.2.2. The original proportion of breathy to modal phonation was always preserved. F0 was normalized for each speaker using the PSOLA (pitch-synchronous overlap and add) function of Praat software. Talker 1 was normalized to a falling f0 of 160-140 Hz, Talker 2 to a falling f0 of 230-240 Hz, and Talker 3 to a falling f0 of 225-217 Hz.

4.2.2.3 Phonetic measurements of the stimuli

CPP, H1-H2, H1-A1, H1-A2, H1-A3, the average of H1-H2 compared to A1, A2-A3, and H2-H4 were measured for all of the stimuli following the same procedure explained in section 3.2.3.

4.2.2.3.1 Results

The results of the acoustic measures for each talker are presented in Figures 48-63. The even numbered figures are graphs of the results of the acoustic measures for each of the 16 phonations, while the odd numbered figures are of the average values per talker. In the odd numbered figures, an asterisk marks the measures that were successful for each talker. A measure was considered successful if there was a significant difference between the phonations. (Significant differences were based on two-tailed *t*-tests.)

Figures 48 and 49 are graphs of the CPP values for the 16 stimuli, and the average per talker, respectively. In both figures, the height of the cepstral peak is plotted along the y-axis in dB. Talkers are given on the x-axis. The lower the cepstral peak, the more breathy the stimuli. The arrow is pointing in the direction of increased breathiness.

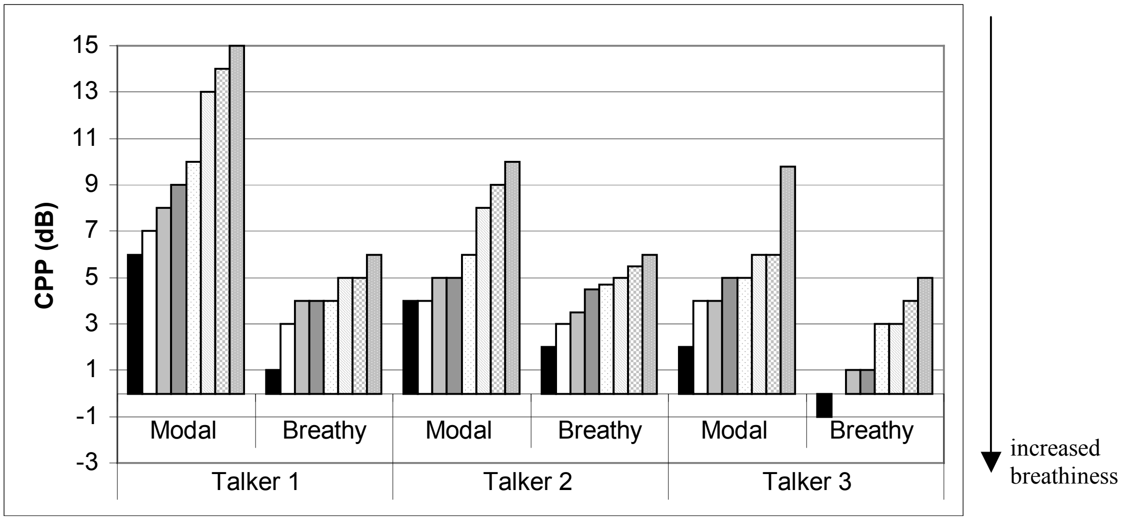


Figure 48: Graph of individual CPP values for three Mazatec talkers.

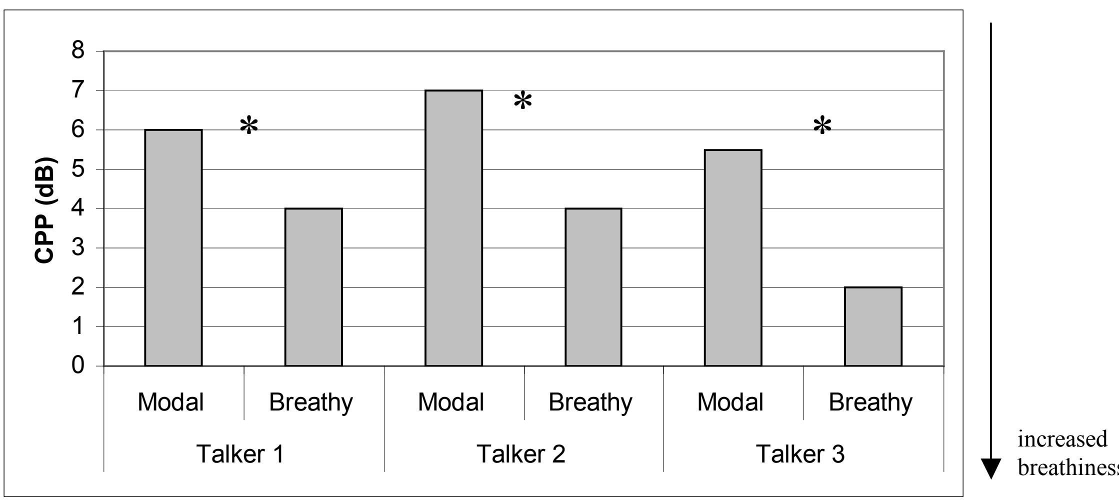


Figure 49: Graph of average CPP values for three Mazatec talkers.

In Figures 50-63, the difference between the amplitudes of the harmonics (H1-H2, H1-A3, H1-A1, H1-A2, the average of H1-H2 compared to A1, A2-A3, and H2-H4) is given on the y-axis in dB. Talkers are given on the x-axis. The greater the difference between the amplitude of the harmonics, the more breathy the stimuli. In each figure, an arrow is pointing in the direction of increased breathiness.

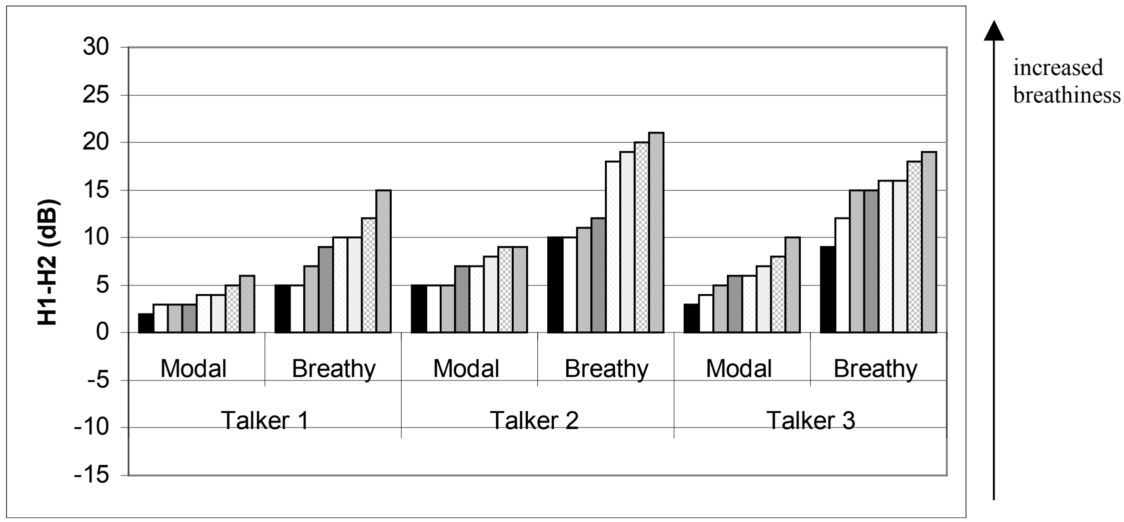


Figure 50: Graph of individual H1-H2 values for three Mazatec talkers.

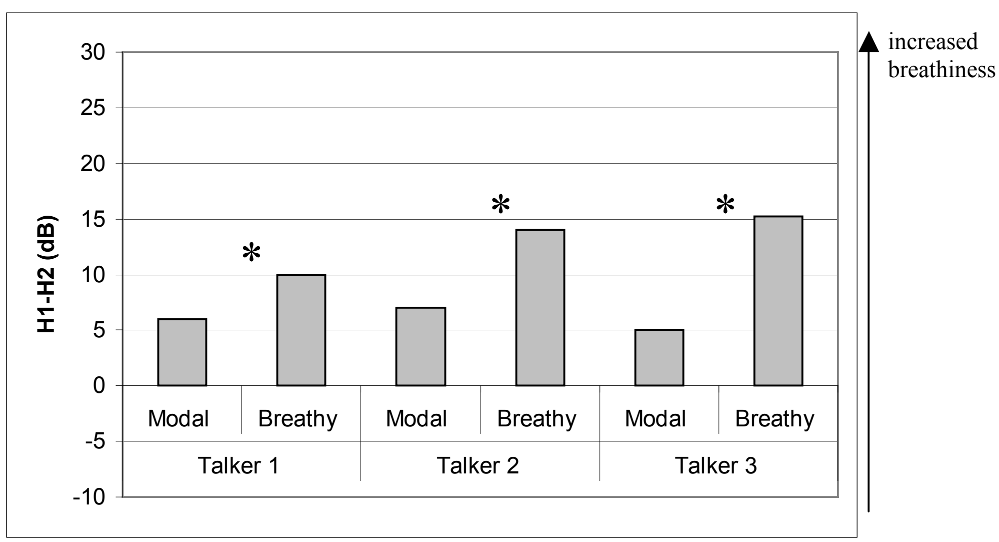


Figure 51: Graph of average H1-H2 for three Mazatec talkers.

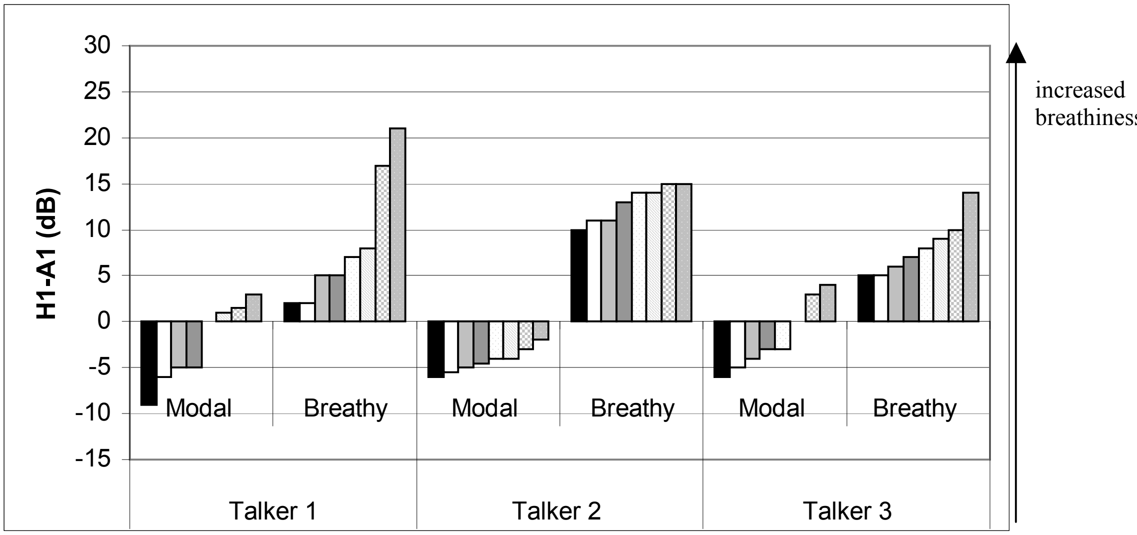


Figure 52: Graph of individual H1-A1 values for three Mazatec talkers.

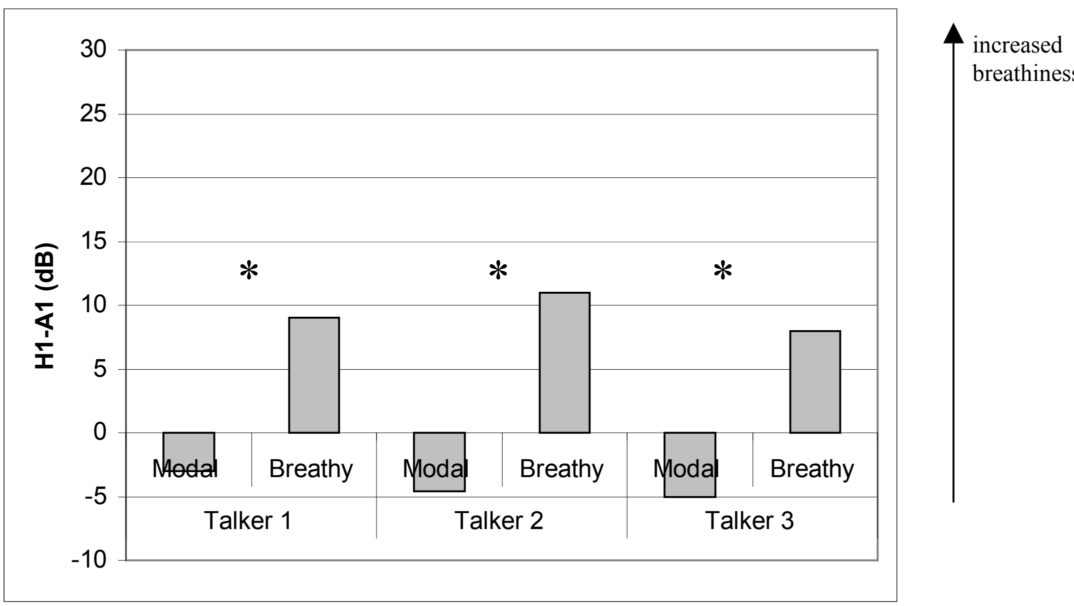


Figure 53: Graph of average H1-A1 values for three Mazatec talkers.

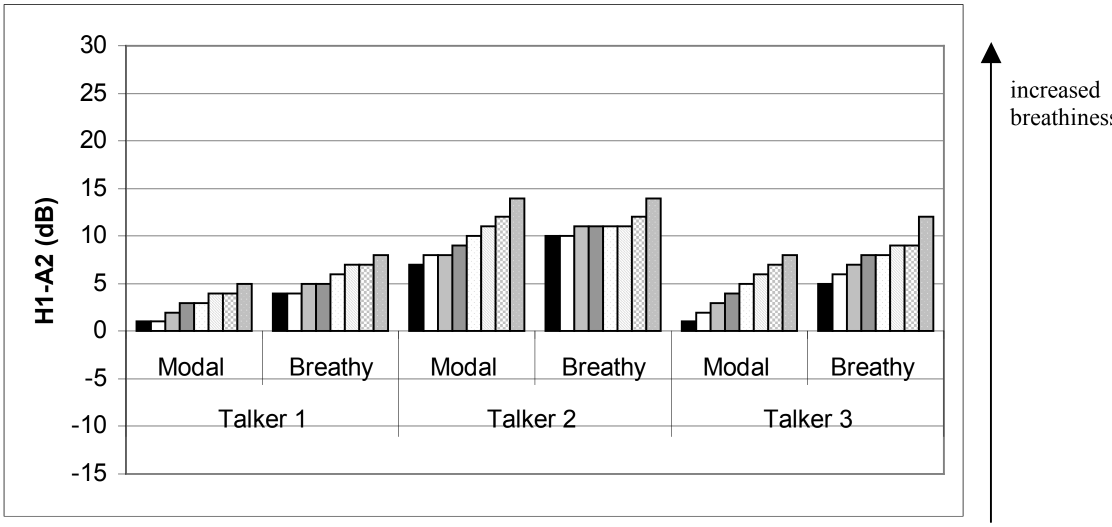


Figure 54: Graph of individual H1-A2 values for three Mazatec talkers.

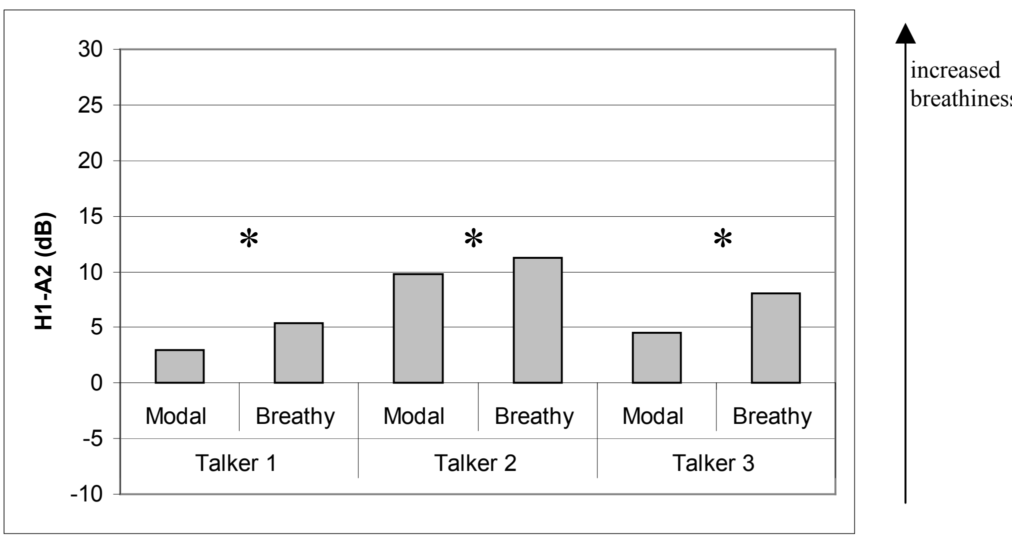


Figure 55: Graph of average H1-A2 values for three Mazatec talkers.

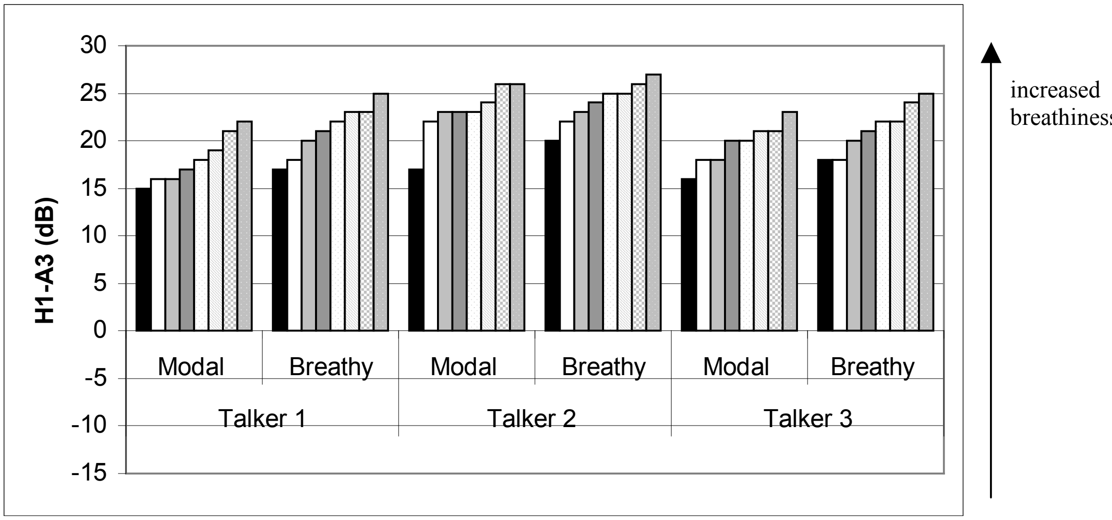


Figure 56: Graph of individual H1-A3 values for three Mazatec talkers.

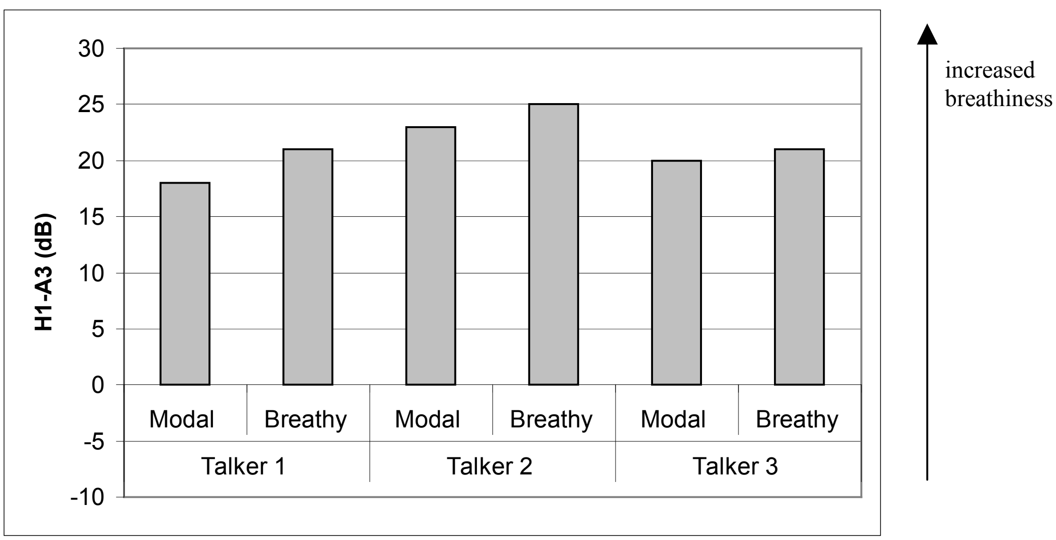


Figure 57: Graph of average H1-A3 values for three Mazatec talkers.

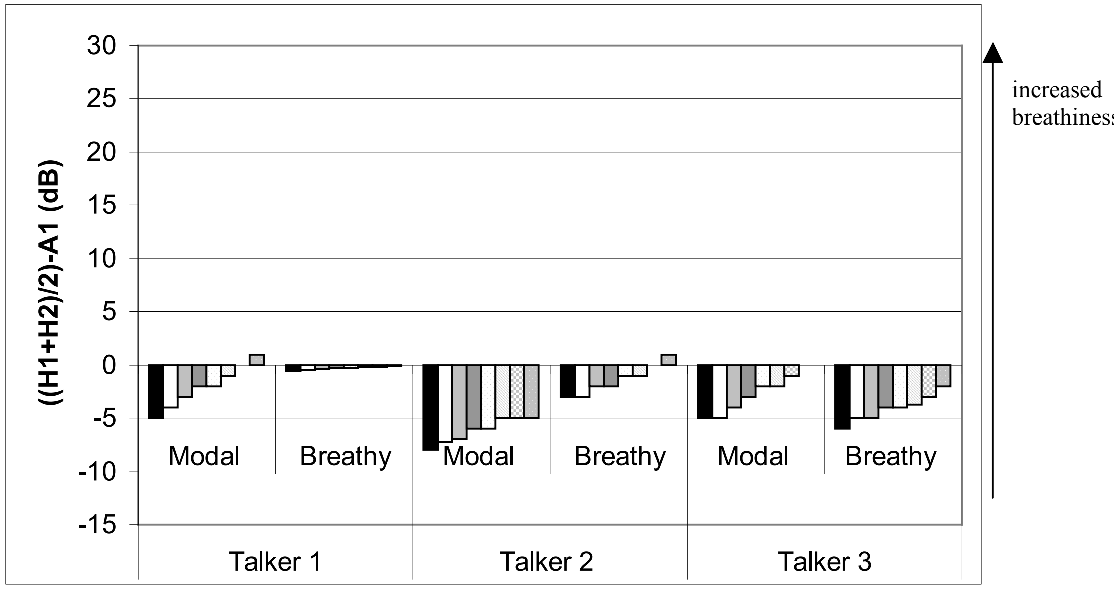


Figure 58: Graph of individual ((H1+H2)/2) – A1 values for three Mazatec talkers.

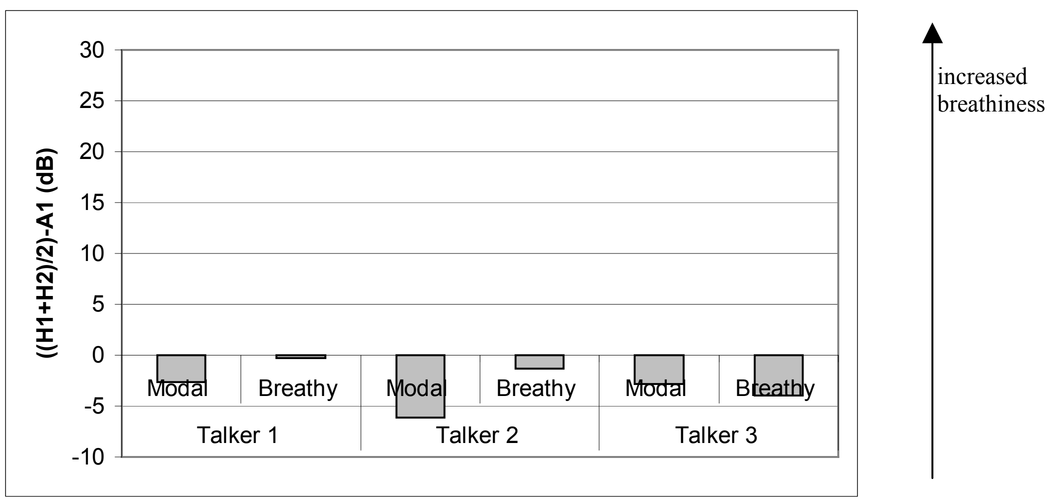


Figure 59: Graph of average ((H1+H2)/2) – A1 values for three Mazatec talkers.

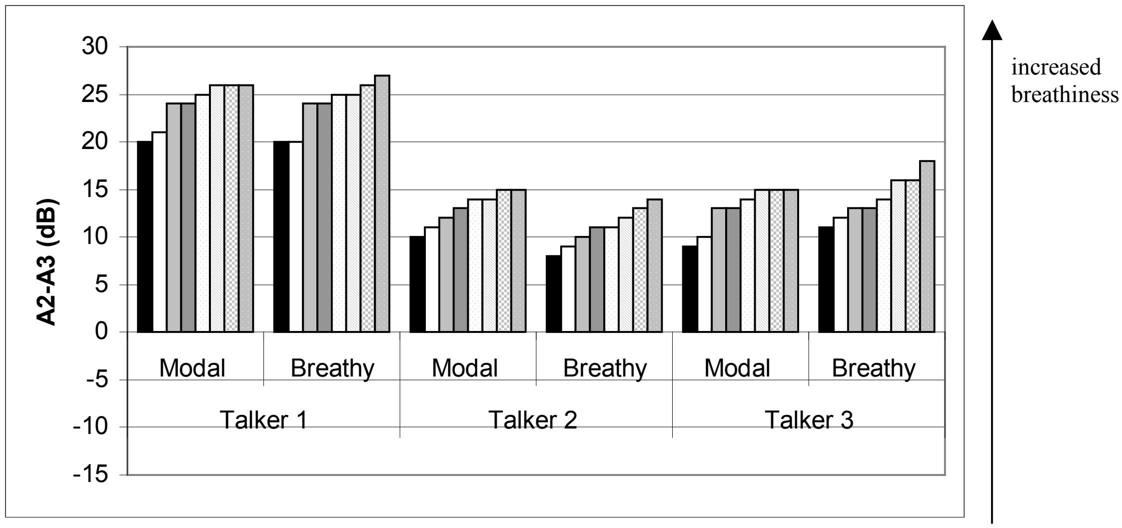


Figure 60: Graph of individual A2-A3 values for three Mazatec talkers.

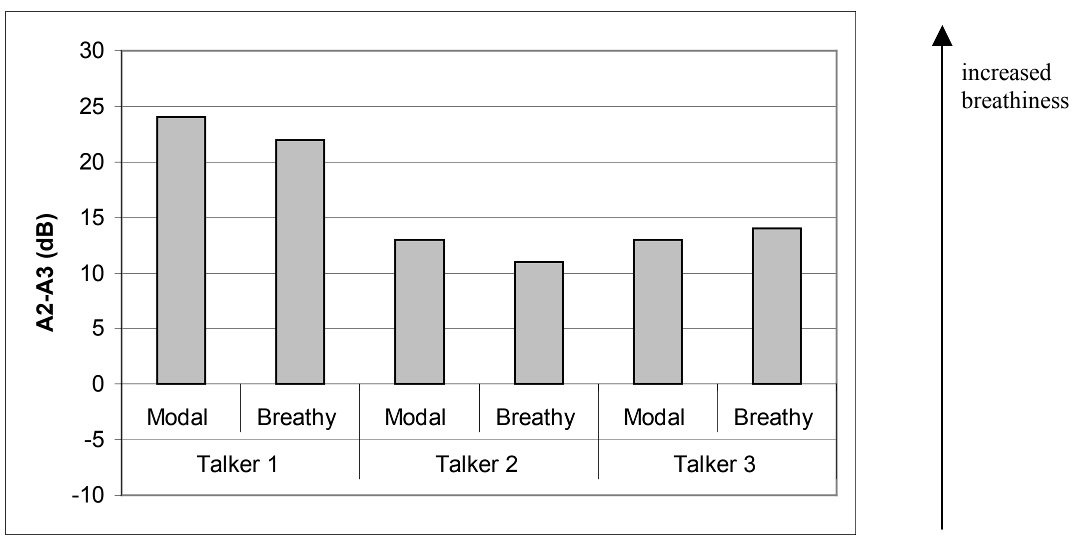
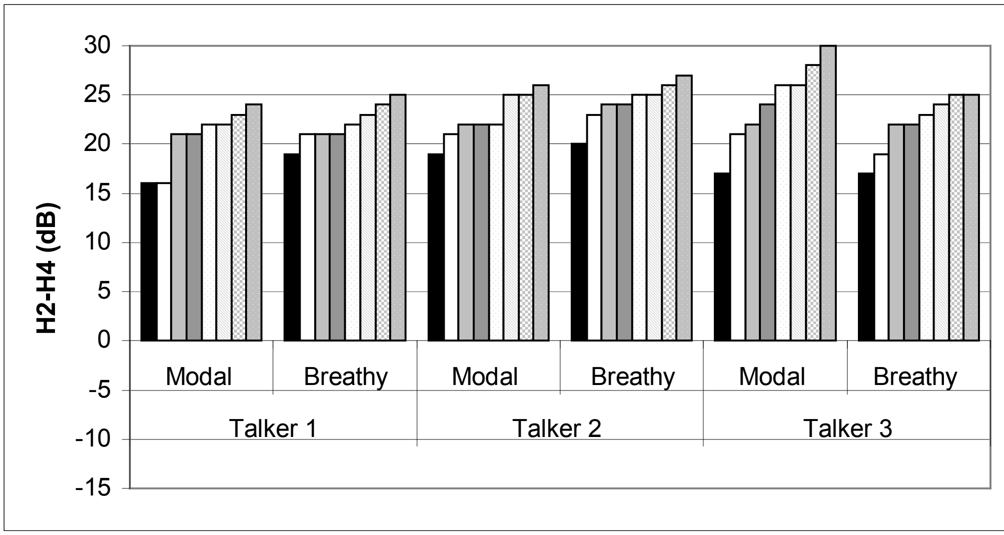
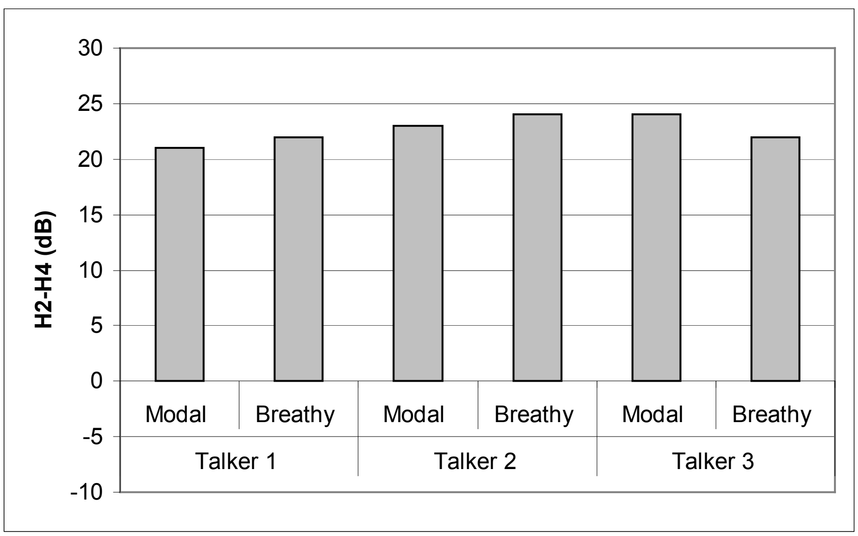


Figure 61: Graph of average A2-A3 values for three Mazatec talkers.



increased
breathiness

Figure 62: Graph of individual H2-H4 values for three Mazatec talkers.



increased
breathiness

Figure 63: Graph of average H2-H4 values for three Mazatec talkers.

Table 6 summarizes the information from the graphs presented in Figures 48-63. A check mark indicates the measure that was successful for each talker; a dashed line indicates that the measure was unsuccessful. A measure was considered successful if there was a significant difference between the phonations in any direction. (Significant differences were based on two-tailed *t-tests*.)

Talkers	Measures							
	CPP	H1-H2	((H1+H2)/2) -A1	H1-A1	H1-A2	H1-A3	A2-A3	H2-H4
Talker 1 (male)	✓	✓	--	✓	✓	--	--	--
Talker 2 (female)	✓	✓	--	✓	✓	--	--	--
Talker 3 (female)	✓	✓	--	✓	✓	--	--	--

Table 6: Success of the measures CPP, H1-H2, ((H1+H2)/2) -A1, H1-A1, H1-A2, H1-A3, H2-H4, A2-A3 for three Mazatec talkers.

From the table presented above, we can see that CPP, H1-H2, H1-A1, and H1-A2 are good measures of phonation for all three talkers. This suggests that for all three talkers, noise, changes in the open quotient, and differences in the speed of vocal fold closure are used to produce phonation contrasts.

Furthermore, on almost all of the measures, except ((H1+H2)/2) -A1, A2-A3, and H2-H4, the values of the spectral measures were very similar for the three talkers. However, on average, on the measures ((H1+H2)/2) -A1, A2-A3, and H2-H4, the values for Talker 3 were in the opposite direction than the values for Talker 1 and Talker 2.

As with the stimuli in the free-sort task, discriminant analysis was performed to determine whether or not these eight measures were sufficient to categorize phonation for these three talkers. This was done by creating two groups, one consisting of the breathy stimuli from all three talkers, and the other consisting of the modal stimuli from all three talkers. The results of the discriminant analysis showed that that these eight acoustic measures accounted for 87% of the variance in the data, with H1-A2 accounting for most of the variance (53% of the variance), followed by H1-A1 (20% of the variance) and then H1-H2 (14% of the variance). The Mazatec stimuli are thus different from the multi-language stimuli used in the previous experiment in that H1-A2 accounts for most of the variance in the data, and H1-H2 accounts for relatively little. Recall that, in the previous experiment, CPP accounting for most of the variance, followed by H1-H2, H1-A2, and then H1-A3.

4.3 Procedure

Each listener participated in three separate similarity-rating tasks, one for each of the three Mazatec talkers. These individual tasks will be referred to as Experiment 1 (which used stimuli from Mazatec Talker 1), Experiment 2 (which used stimuli from Mazatec Talker 2), and Experiment 3 (which used stimuli from Mazatec Talker 3) throughout. The similarity-rating tasks were randomly ordered with respect to each other and with respect to the task presented in Chapter 3. Each similarity-rating task was implemented in a program created by the Bureau of Glottal Affairs, Division of Head/Neck Surgery, UCLA School of Medicine and was run on a laptop with Sennheiser HD 280 professional headphones and a Telex p500 external soundcard. Stimuli were presented in pairs separated by .5 seconds. The vowels in each pair were produced by the same talker and were matched for vowel quality. Listeners heard both orders of the stimuli (AB and BA), but not the stimulus paired with itself (AA or BB). The vowel that occurred first within the pair (AB or BA) was chosen randomly. 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.

Listeners began with a practice session composed of one of the Mazatec talkers not being used in the experiment. Again, the directions were given in the listeners' native language, following the same procedure explained in Chapter 3. 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” (see Figure 64). 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. Due to the large number of stimulus pairs (240 pairs per talker), listeners were instructed to work as quickly as they could, but not at an uncomfortable pace.

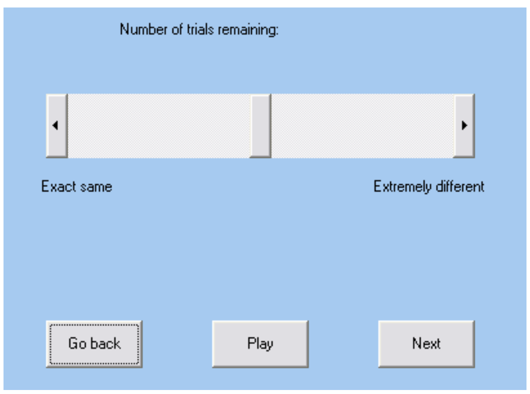


Figure 64: Screenshot of the similarity-rating task.

The results were analyzed following the procedures using in numerous studies on the perception of pathologically-disordered phonation (e.g. Kreiman et al. 1990, Kreiman et al. 1994 and 1996, 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. For each experiment, 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 (roughly equivalent to the scale used in Kreiman et al. 1990) or more for a single listener's rating of AB and BA pairs and 2) calculating the correlation (using Pearson's r) 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 multidimensional scaling procedure known as 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. For each of the three similarity-rating tasks, separate INDSCAL solutions in two 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 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.

4.4 Results

In section 4.4.1, the within-listener agreement for all three experiments will be presented by listeners' language. In section 4.4.2, separate multidimensional scaling solutions will be presented for all three experiments.

4.4.1 Within-listener agreement

The results of the average within-listener agreement are presented by language in Tables 7-9. The range of values is also presented. The percent of responses for a single listener's rating of the AB and BA order of a pair of voices that differed by a value of 40 or more is labeled "% ratings > +/- 40 scale value" and the correlation between the first and the second rating of each pair of stimuli by a single listener is labeled "r for 1st and 2nd ratings" in all three tables.

English listeners				
	% ratings > +/- 40 scale value		r for 1st and 2nd ratings	
Experiment	**mean**	**range**	**mean**	**range**
Experiment 1	10.5%	1.7-15.3%	.71	.50-.72
Experiment 2	8.6%	1.9-18.5%	.72	.55-.74
Experiment 3	10.0%	0-22%	.73	.51-.74

Table 7: Within-listener agreement for English listeners.

Spanish listeners				
	% ratings > +/- 40 scale value		r for 1st and 2nd ratings	
Experiment	**mean**	**range**	**mean**	**range**
Experiment 1	11.9%	.8-23.0%	.73	.47-.72
Experiment 2	8.8%	3.7- 22.2%	.72	.50-.74
Experiment 3	12.0%	1.9-20.0%	.75	.48-.75

Table 8: Within-listener agreement for Spanish listeners.

Gujarati listeners				
	% ratings > +/- 40 scale value		r for 1st and 2nd ratings	
Experiment	**mean**	**range**	**mean**	**range**
Experiment 1	3.8%	.8-4.5%	.90	.75-.95
Experiment 2	4.5%	1.9-5.1%	.91	.75-.93
Experiment 3	4.2%	1.9-5.23%	.89	.77-.95

Table 9: Within-listener agreement for Gujarati listeners.

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 for Experiment 1, 8.6% for Experiment 2, and 10.0% for Experiment 3. In addition, the rating and re-rating of pairs (that is, AB and BA) were correlated for the English listeners ($r = .71$ in Experiment 1, $r = .72$ in Experiment 2, and $r = .73$ in Experiment 3).

On average, Spanish listeners deviated by more than a scale value of 40 on only 11.9% of the pairs for Experiment 1, 8.8% for Experiment 2, and 12.0% for Experiment 3. Once again, we see that Spanish listeners were behaving very similarly to the English listeners, despite the allophonic breathiness in English. In addition, the rating and re-rating of pairs (that is, AB and BA) were correlated for the Spanish listeners ($r = .73$ in Experiment 1, $r = .72$ in Experiment 2, and $r = .75$ in Experiment 3).

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 for Experiment 1, 4.5% for Experiment 2, and 4.2% for Experiment 3. In addition, the rating and re-rating of pairs (that is, AB and BA) were well correlated for the Gujarati listeners ($r = .90$ in Experiment 1, $r = .91$ in Experiment 2, and $r = .89$ in Experiment 3).

4.4.2 Multidimensional scaling

As previously mentioned, listeners' responses were analyzed using the INDSCAL model of MDS. Recall that separate solutions in one to five dimensions were found for each listener group, and that R^2 and stress were used to find the location of the elbow, which allows you to give a best estimate of the number of dimensions the listeners used to make their judgments. In the next section, both group and individual MDS solutions will be presented for each experiment.

4.4.2.1 Experiment 1 (Talker 1)

4.4.2.1.3 Multidimensional scaling analyses

Figures 65-67 show the R^2 and stress values for one to five dimensions for English, Spanish, and Gujarati listeners, respectively. The number of dimensions is plotted on the x-axis. The R^2 and stress values are plotted on the y-axis in an unspecified value. In each graph, an arrow is pointing to the elbow.

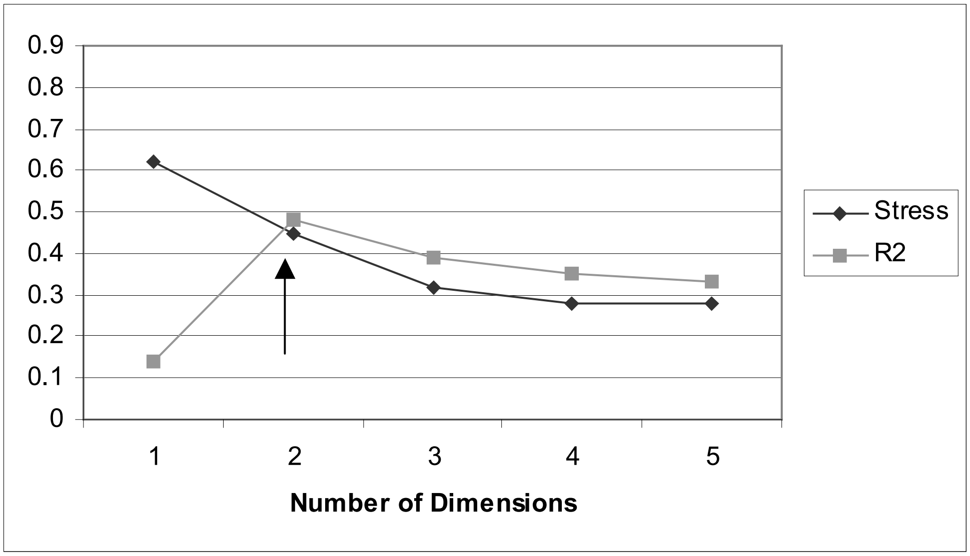


Figure 65: Graph of R^2 and stress for English listeners (Experiment 1).

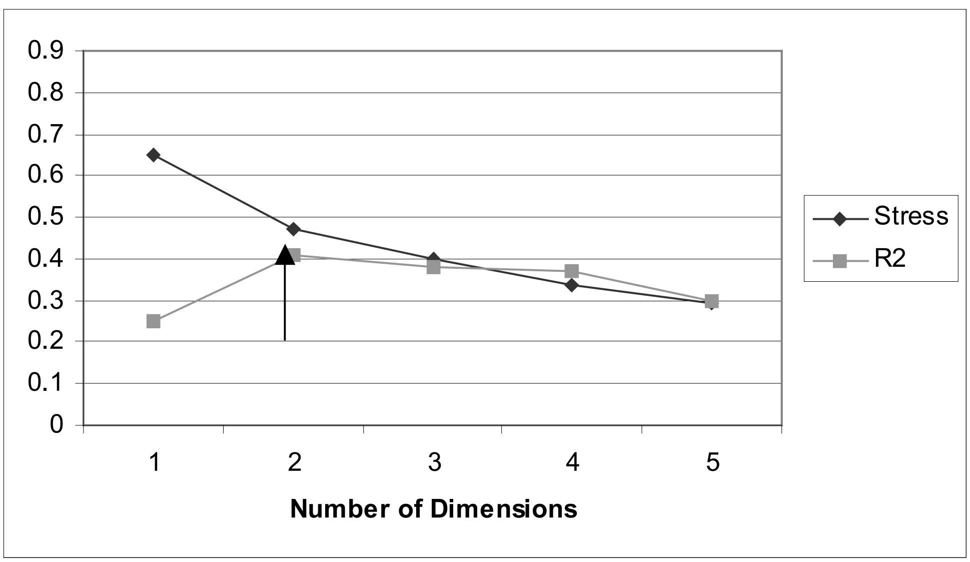


Figure 66: Graph of R^2 and stress for Spanish listeners (Experiment 1).

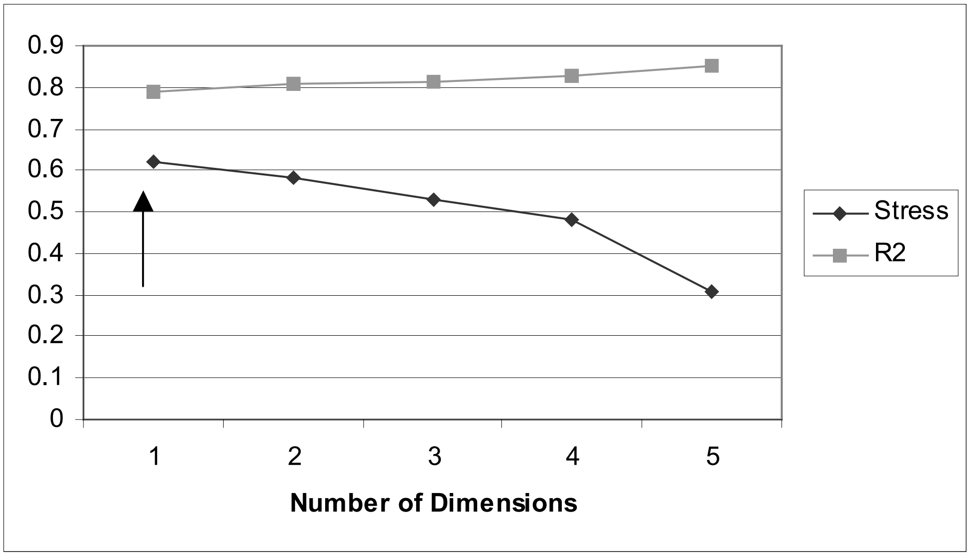


Figure 67: Graph of R^2 and stress for Gujarati listeners (Experiment 1).

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 .45), while a three-dimensional solution accounted for 49% of the variance (with a stress value of .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 indicated a two-dimensional solution, which accounted for 41% of the variance in the rating data (with a stress value of .46). Additional dimensions actually accounted for less variance. For example, adding a third dimension to the solution accounted for 38% of the variance, a 3% reduction in the amount of variation accounted for by the two-dimensional solution.

For the Gujarati listeners, stress and R^2 values indicated a one-dimensional solution, which accounted for 79% variance (with a stress value of .62). After the first dimension, the amount of variance accounted for increased only gradually. For example, a two-dimensional solution only increased the amount of variance accounted for by 2%.

4.4.2.1.3 Group perceptual space for experiment 1

In the following sections, figures 68, 70, and 72 represent the perceptual space for Experiment 1 for English, Spanish and Gujarati listeners, respectively. For all three graphs, each point represents one of the 16 stimuli in the experiment. Here, and throughout, voices that are perceptually similar are closer together in the space. For some graphs, such as Figure 68, 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.

4.4.2.1.2.1 Perceptual space for the English listener (Experiment 1)

The English listeners' perceptual space is presented in Figure 68, with dimension one graphed on the x-axis, and dimension two on the y-axis.

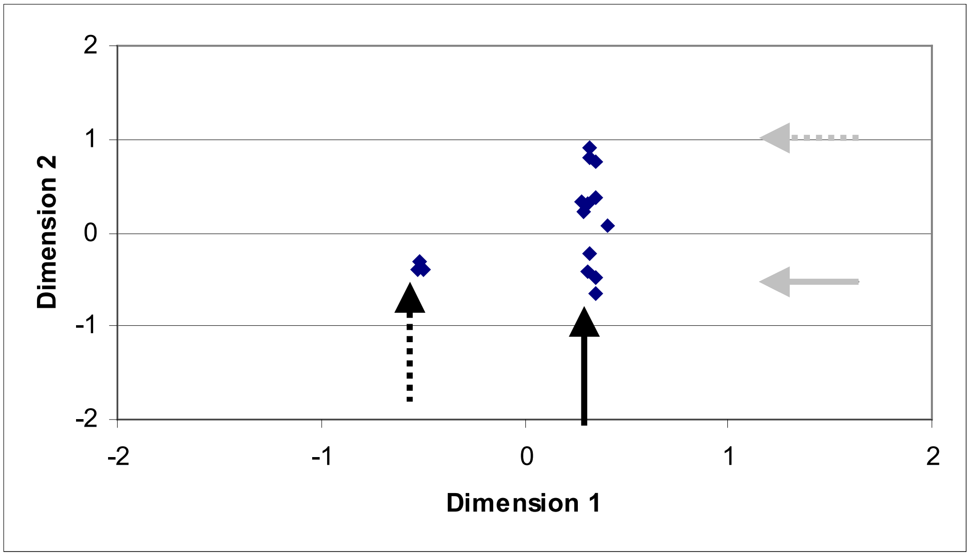


Figure 68: Group perceptual space for English listeners (Experiment 1).

In Figure 68 (English listeners, Talker 1), dimension one represents H1-H2 (there was a significant correction between dimension one and H1-H2 ($r = .83$, $p < .05$)), with the stimuli mapped into two groups. The four leftmost points (indicated by the dotted black arrow) correspond to stimuli with H1-H2 values of approximately 10-15 dB, and the rightmost 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, 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 (that is, 6-10 dB). There was a significant correlation between dimension two and CPP ($r = .85$, $p < .05$). Furthermore, the mapping of the stimuli along dimension one was categorical, while the mapping of the stimuli along dimension two was more continuous.

One could ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 68. Figure 69 is a scatter plot of the measured H1-H2 and CPP values of the stimuli.

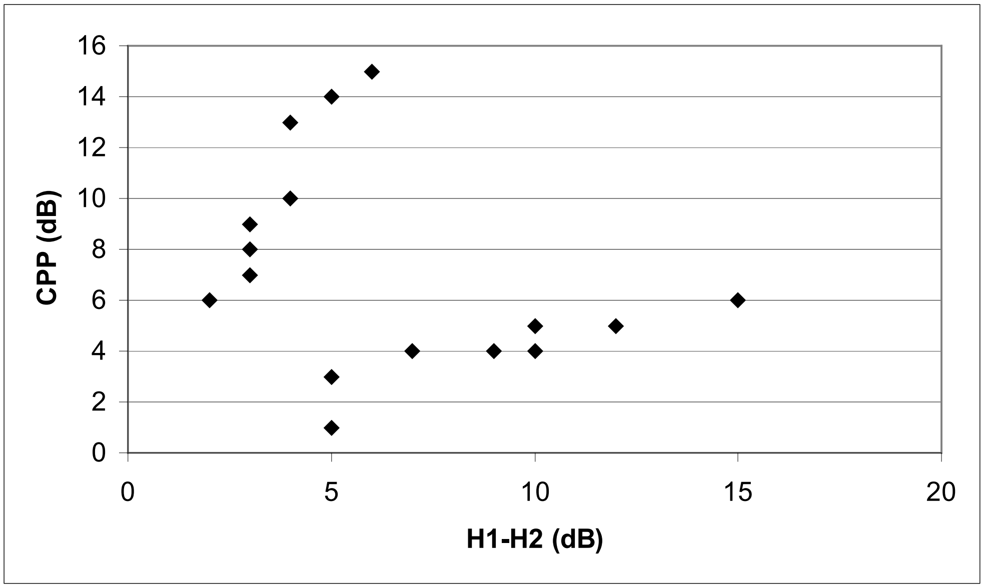


Figure 69: Graph of the measured H1-H2 and CPP values of the stimuli from Experiment 1.

In Figure 69, 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 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 Figure 68. However, there are no gaps in the stimuli that correspond to any of the other divisions that the English listeners made along the CPP dimension.

4.4.2.1.2.3 Perceptual space for the Spanish listeners (Experiment 1)

The Spanish listeners' perceptual space is presented in Figure 70, with dimension one graphed on the x-axis, and dimension two on the y-axis.

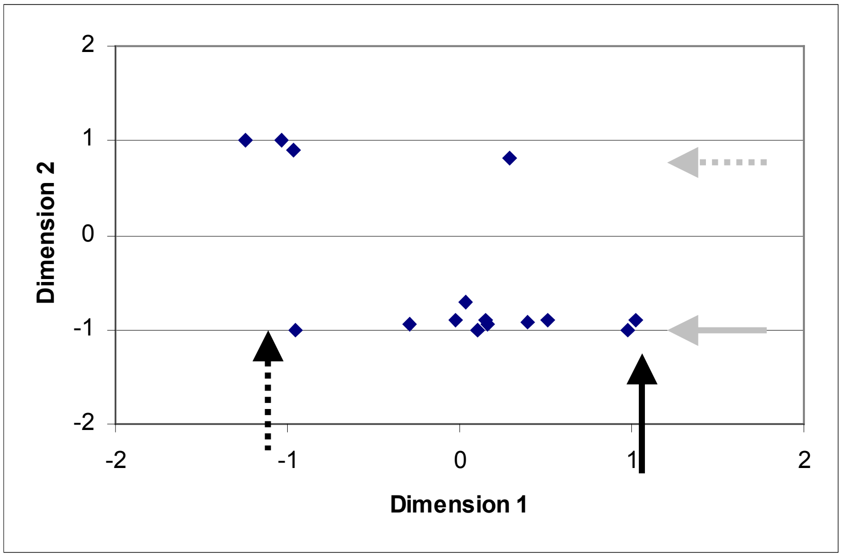


Figure 70: Group perceptual space for Spanish listeners (Experiment 1).

Dimension one represents H1-A1, with the four leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-A1 values of approximately –5 to –9 dB, and the two rightmost 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 (that is, 0-8 dB). There was a significant correlation between dimension one and H1-A1 ($r = .79$, $p < .05$). Dimension two represents H1-H2 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. There was a significant correlation between dimension two and H1-H2 ($r = .84$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 70. Figure 71 is a scatter plot of the measured H1-A1 and H1-H2 values of the stimuli.

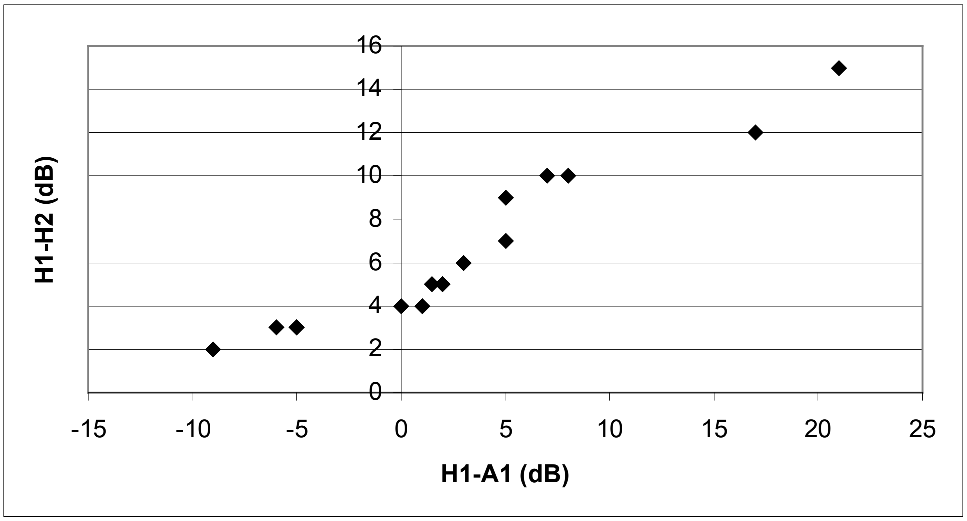


Figure 71: Graph of the measured H1-A1 and H1-H2 values of the stimuli from Experiment 1.

From Figure 71, we can see that, along both dimensions, there are some gaps between the stimuli. For both dimensions, Spanish listeners' judgments corresponded to one of the 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 Figure 70. However, along the H1-H2 dimension, there is no gap that corresponds directly to the Spanish listeners' division in Figure 70.

4.4.2.1.2.3 Perceptual space for Gujarati listeners (Experiment 1)

For the Gujarati results (Figure 72), there is only one dimension, which is plotted along the x-axis. However, to make the perceptual space easier to represent graphically, a second "dimension" was added to the graph along the y-axis. Note how the values along dimension 2 are all zero.

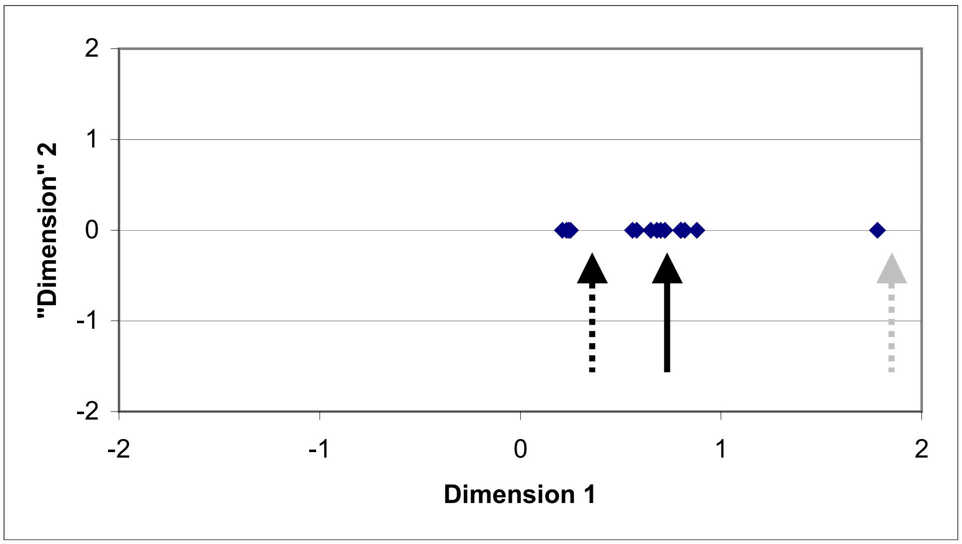


Figure 72: Group perceptual space for Gujarati listeners (Experiment 1).

In Figure 72, the stimuli cluster into three groups, distinguished by dimension 1, which represents H1-H2. The leftmost 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 between 4 and 12 dB, and the rightmost point (indicated by the solid gray arrow) corresponds to the stimulus with a H1-H2 value of 15 dB. There was a significant correlation between dimension one and H1-H2 ($r = .94$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 72. Figure 73 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 Figure 72.

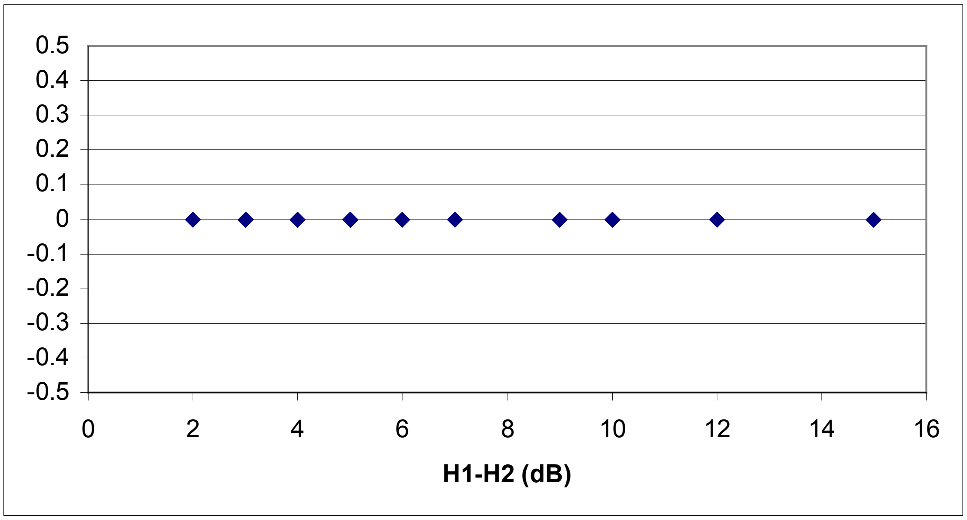


Figure 73: Graph of the measured H1-H2 values of the stimuli from Experiment 1. (A second “dimension” was added to the graph along the y-axis to parallel Figure 72.)

From Figure 73, we can see that 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 (Figure 72). 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 (Figure 72).

4.4.2.1.3 Differences among individual listeners

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 are accounted for in INDSCAL.) The higher the value of the weight, the more important the dimension is to that particular listener.

Tables 10, 11, and 12 are of the individual subject weights for English, Spanish, and Gujarati listeners, respectively. The average value of the weight for each dimension is given at the end of each table.

English listeners		
Listeners	**Dimension 1 (H1-H2)**	**Dimension 2 (CPP)**
1	0.43	0.32
2	0.36	0.33
3	0.45	0.33
4	0.32	0.25
5	0.4	0.35
6	0.44	0.42
7	0.36	0.26
8	0.46	0.34
9	0.43	0.35
10	0.43	0.34
11	0.37	0.31
12	0.43	0.32
13	0.43	0.35
14	0.53	0.42
15	0.34	0.31
16	0.34	0.31
17	0.4	0.31
18	0.55	0.43
Average	**.42**	**.33**

Table 10: Importance of individual dimensions to English listeners.

Spanish listeners		
Listeners	**Dimension 1 (H1-A1)**	**Dimension 2 (H1-H2)**
1	0.33	0.29
2	0.44	0.42
3	0.35	0.24
4	0.4	0.31
5	0.43	0.41
6	0.46	0.42
7	0.45	0.34
8	0.57	0.37
9	0.39	0.34
10	0.48	0.36
11	0.43	0.36
12	0.43	0.35
13	0.55	0.4
14	0.56	0.41
15	0.61	0.35
16	0.68	0.31
17	0.57	0.25
18	0.68	0.3
Average	**.49**	**.35**

Table 11: Importance of individual dimensions to Spanish listeners.

Gujarati listeners	
Listeners	**Dimension 1 (H1-H2)**
1	0.89
2	0.78
3	0.81
4	0.75
5	0.9
6	0.88
7	0.87
8	0.85
9	0.82
10	0.78
11	0.79
12	0.73
Average	**.82**

Table 12: Importance of individual dimensions to Gujarati listeners.

For English, results showed that listeners differed in how they weighted each dimension. However, the English listeners were consistent in that all listeners agreed on the relative order of importance of the dimensions (Table 10). For all 18 listeners considered individually, the first dimension of the group solution (H1-H2) was always more important than the second of the group solution (CPP). That is, each listener weighed dimension one for the group solution as his or her own individual dimension one.

For Spanish listeners, results showed that listeners also differed in how they weighted each cue. However, as with the English listeners, the Spanish listeners agreed on the relative order of importance of the dimensions (Table 11). For all 18 listeners considered individually, the first dimension (H1-A1) was always more important than the second (H1-H2). That is, each listener weighed dimension one for the group solution as his or her own individual dimension one.

For the Gujarati listeners, we cannot say anything about the relative importance of multiple dimensions beyond that fact that they only used one dimension (i.e., it is all or nothing for the Gujarati listeners). Therefore, individual MDS analyses were run on the Gujarati listeners to see if each listener used H1-H2 as his/her most important dimension. Results showed that the Gujarati listeners were consistent in that all of the listeners considered individually had a one-dimensional perceptual space, and this one dimension correlated with H1-H2 for all 12 listeners.

4.4.2.2 Experiment 2 (Talker 2)

4.4.2.2.1 Multidimensional scaling analyses

Figures 74-76 show the R^2 and stress values for a one to five dimension solution for English, Spanish, and Gujarati listeners, respectively. In each graph, an arrow is pointing to the elbow.

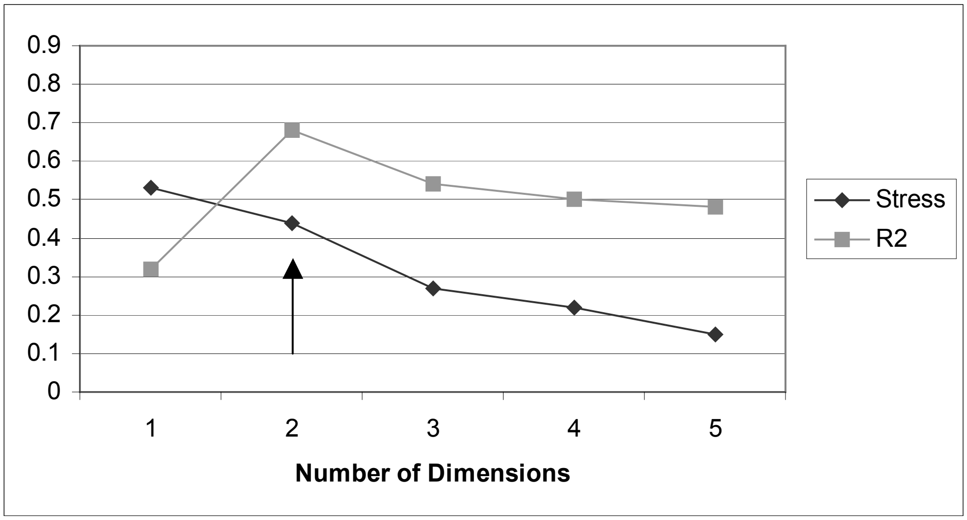


Figure 74: Graph of R^2 and stress for English listeners.

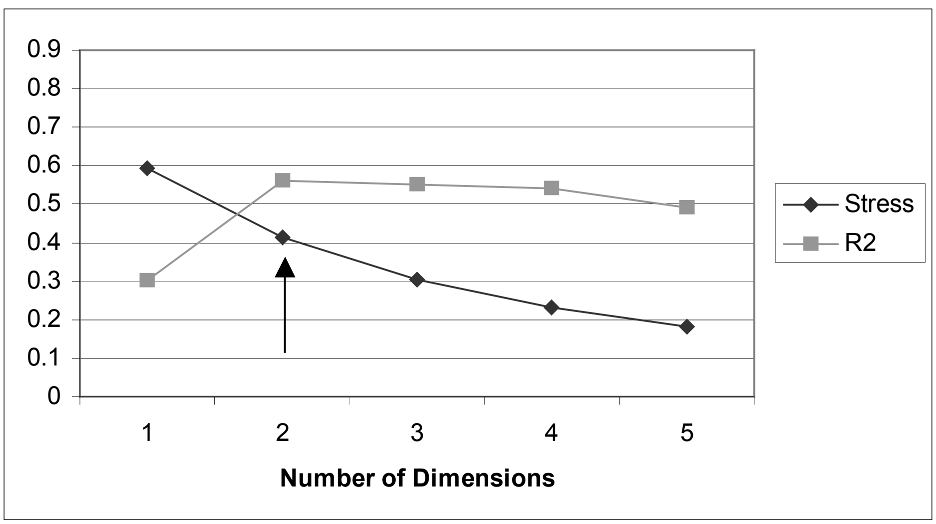


Figure 75: Graph of R^2 and stress for Spanish listeners.

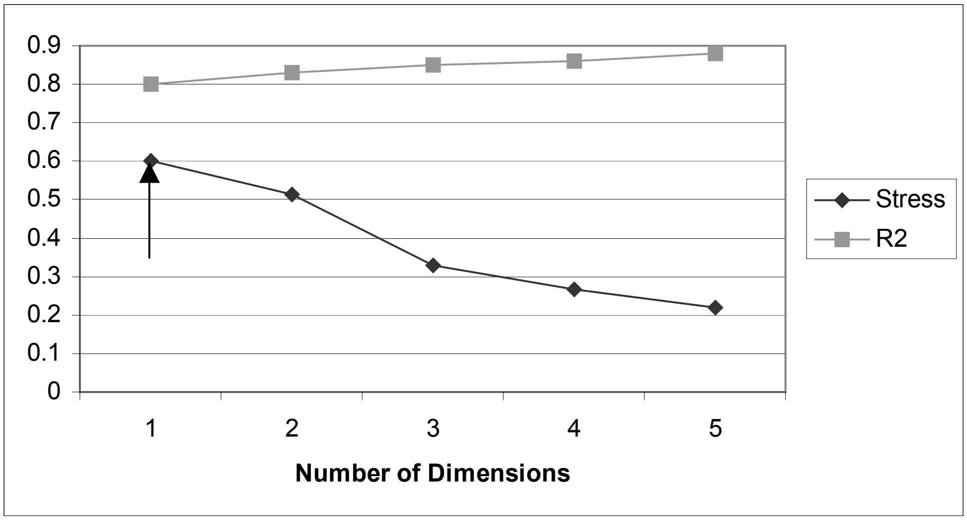


Figure 76: Graph of R^2 and stress for Gujarati listeners.

For the English listeners, stress and R^2 values indicated a two-dimensional solution, which accounted for 68% of the variance (with a stress value of .44) in the rating data. Additional dimensions accounted for less variance, with a third dimension accounting for only 54% of the variance.

For the Spanish listeners, stress and R^2 values also indicated a two-dimensional solution, which accounted for 56% of the variance (with a stress value of .41) in the rating data. As with the English listeners, additional dimensions accounted for less variance.

For the Gujarati listeners, stress and R^2 values indicated a one-dimensional solution, which accounted for 80% variance (with a stress value of .60). After the first dimension, the amount of variance accounted for increased only gradually.

4.4.2.2.2 Group perceptual space

Figures 77, 79, and 81 represent the perceptual space in Experiment 2 for English, Spanish and Gujarati, respectively. The dimensions were interpreted using the same procedure explained in Experiment 1 (section 4.4.2.1.2).

4.4.2.2.2.1 Perceptual space for English listeners (Experiment 2)

The English listeners' perceptual space is presented in Figure 77, with dimension one graphed on the x-axis, and dimension two on the y-axis.

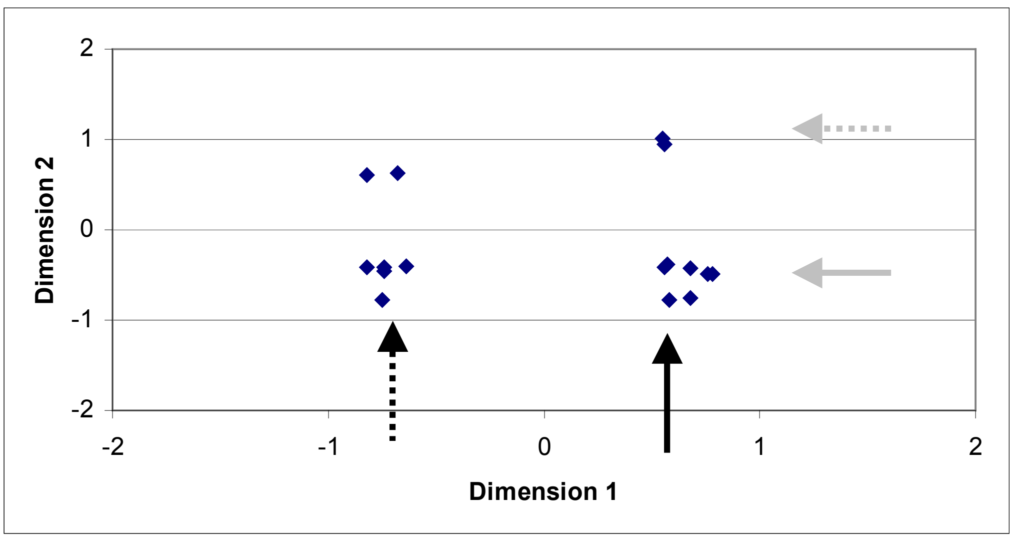


Figure 77: Group perceptual space for English listeners (Experiment 2).

The English listeners organized the perceptual space along two dimensions and divided it into approximately four categories. Dimension one represents H1-H2, with the leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-H2 values of approximately 10-21 dB, and the rightmost points (indicated by the solid black arrow) corresponding to stimuli with H1-H2 values less than or equal to 9 dB. There was a significant correlation between dimension one and H1-H2 ($r = .83$, $p < .05$). Dimension two represents CPP with the topmost points (indicated by the dotted gray arrow) corresponding to stimuli with CPP values of approximately 8 dB or greater. The bottom-most points (indicated by the solid gray arrow) represent stimuli with lower CPP values of less than or equal to 6 dB. Unlike Experiment 1, there are fewer intermediate points along dimension two. The reason for this is that the stimuli in Experiment 2 did not have the wide range of CPP values that the stimuli in Experiment 1 had. (Compare Talker 1 and Talker 2's CPP values in Figure 48. For Talker 2, the source of the stimuli used in Experiment 2, most CPP values are below five dB. This explains why all the stimuli in Figure 74 are grouped together along the bottom half of dimension two.) There was a significant correlation between dimension two and CPP ($r = .81$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 77. Figure 78 is a scatter plot of the measured H1-H2 and CPP values of Talker 2's stimuli.

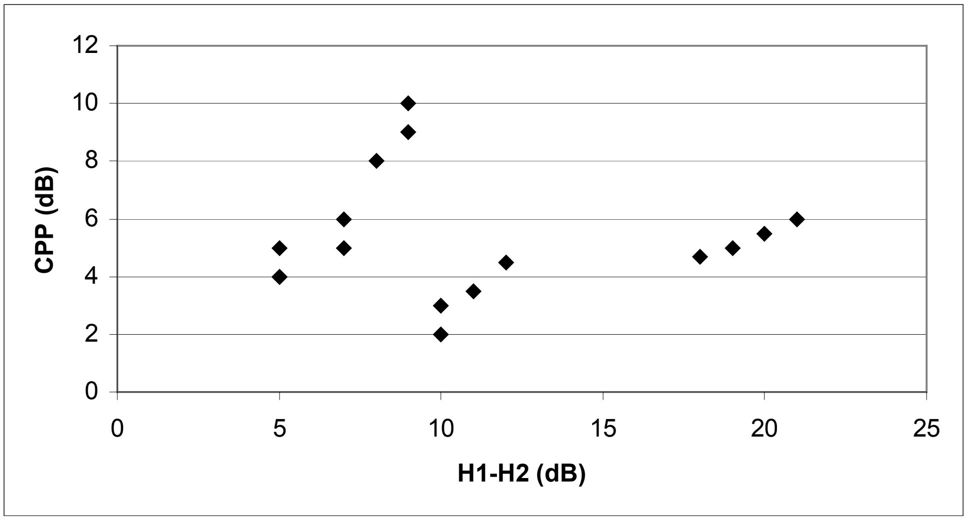


Figure 78: Graph of the measured H1-H2 and CPP values of the stimuli from Experiment 2.

From Figure 78, we can see that, along both dimensions, there are some gaps between the stimuli. However, there is no gap that corresponds to the English listeners' judgments along H1-H2. (There is a larger gap between the stimulus with an H1-H2 value of 12 dB and the stimulus with a H1-H2 value of 18 dB, but the English listeners did not use this gap to make their judgments.) Along the CPP dimension, only one of the gaps corresponds to the English listeners' judgments. There is a gap between the stimulus with a CPP value of 6 dB and the stimulus with a CPP value of 8 dB. This gap corresponds directly to the English listeners' division along CPP in Figure 77.

4.4.2.2.2.2 Perceptual space for Spanish listeners (Experiment 2)

The Spanish listeners' perceptual space is presented in Figure 79, with dimension one graphed on the x-axis, and dimension two on the y-axis.

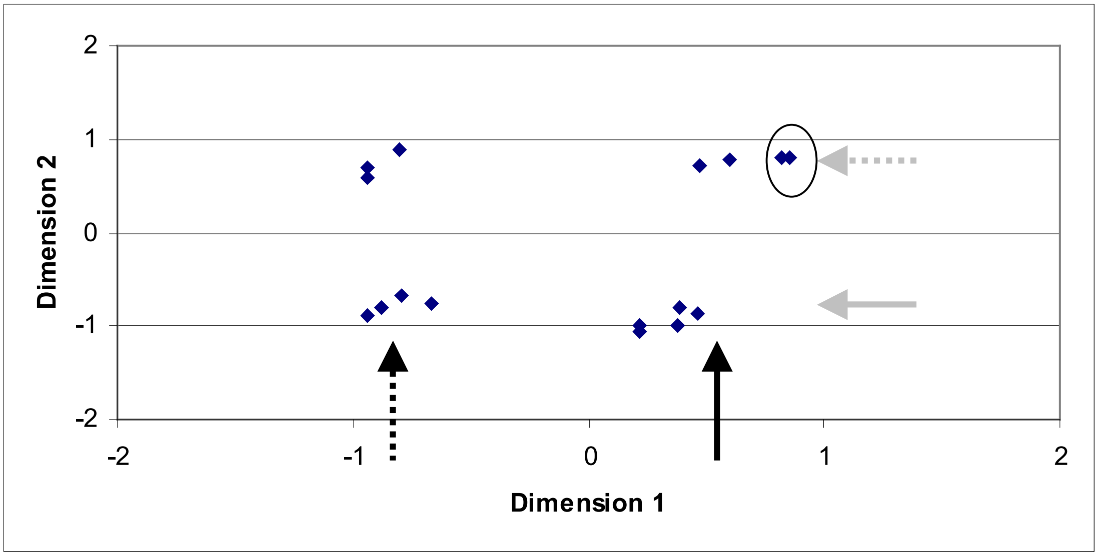


Figure 79: Group perceptual space for Spanish listeners (Experiment 2).

The Spanish listeners also organized the perceptual space along two dimensions and divided it into approximately four categories. Dimension one represents H1-A1. Here the points are divided into two groups, with the leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-A1 values of approximately -6 to -2 dB, and the rightmost points (indicated by the solid black arrow) corresponding to stimuli with H1-A1 values of 10-15 dB. (The two rightmost points, circled in the above figure, represent the two stimuli with H1-A1 values of 15 dB.) There was a significant correlation between dimension one and H1-A1 ($r = .86$, $p < .05$). Dimension two represents H1-H2. The topmost points (indicated by the dotted gray arrow) represent stimuli with H1-H2 values of approximately 9-21 dB, and the bottom-most points (indicated by the gray arrow) corresponding to stimuli with H1-H2 values less than 9 dB. There was a significant correlation between dimension two and H1-H2 ($r = .78$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 79. Figure 80 is a scatter plot of the measured H1-A1 and H1-H2 values of the stimuli.

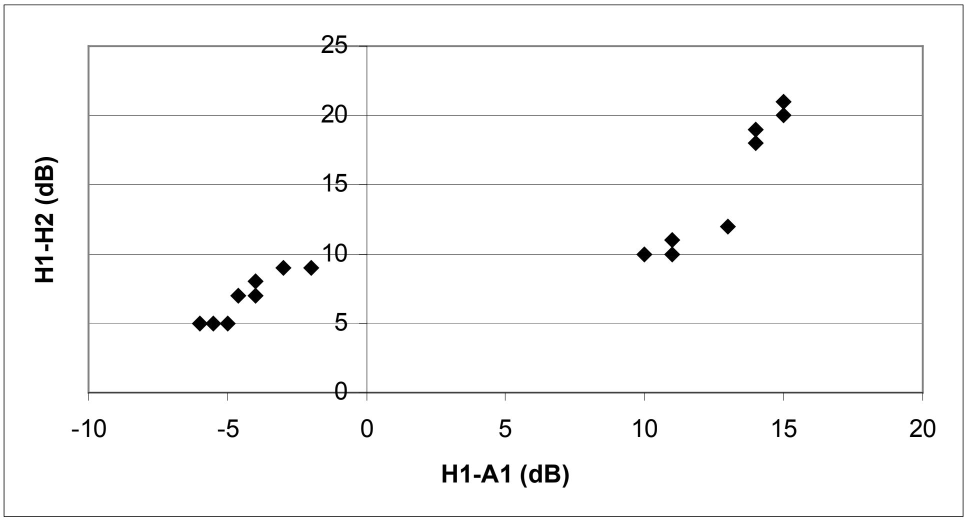


Figure 80: Graph of the measured H1-A1 and H1-H2 values of the stimuli from Experiment 2.

From Figure 80, we can see that, along both dimensions, there are some gaps between the stimuli. Along the H1-A1 dimension, there is a very large gap between the stimulus with an H1-A1 value of -2 dB and the stimulus with an H1-A1 value of 10 dB. This corresponds directly to the Spanish listeners' division along H1-A1 in Figure 79. Along the H1-H2 dimension, there is no gap corresponding to the Spanish listeners' judgments. There is a larger gap between the stimulus with an H1-H2 value of 12 dB and the stimulus with a H1-H2 value of 18 dB, but the Spanish listeners did not use this gap to make their judgments.

4.4.2.2.2.3 Perceptual space for the Gujarati listeners (Experiment 2)

For the Gujarati results (Figure 81) there is only one dimension, which is plotted along the x-axis. However, to make the perceptual space easier to represent graphically, a second "dimension" was added to the graph along the y-axis. Note how the values along dimension 2 are all zero.

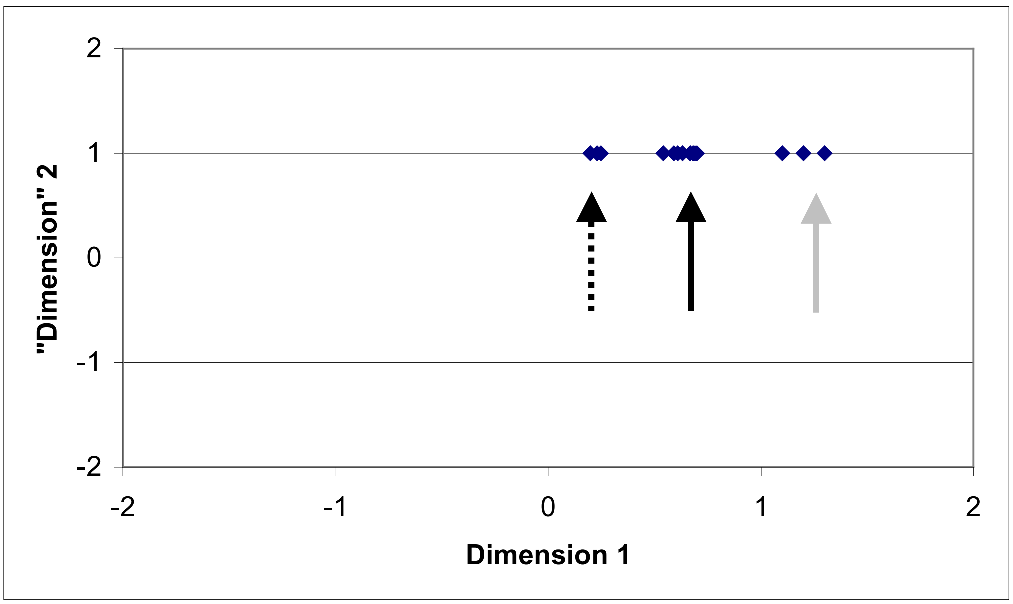


Figure 81: Group perceptual space for Gujarati listeners (Experiment 2).

In Figure 81, the stimuli cluster into approximately three groups, distinguished by dimension one, which represents H1-H2. The leftmost points (indicated by the dotted black arrow) correspond to the three stimuli with H1-H2 values of 5 dB, the middle set of points (indicated by the solid black arrow) correspond to stimuli with H1-H2 between 7-12 dB, and the rightmost points (indicated by the solid gray arrow) correspond to stimuli with H1-H2 values of 18-21 dB. There was a significant correlation between dimension one and H1-H2 ($r = .90$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 81. Figure 82 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 Figure 81.

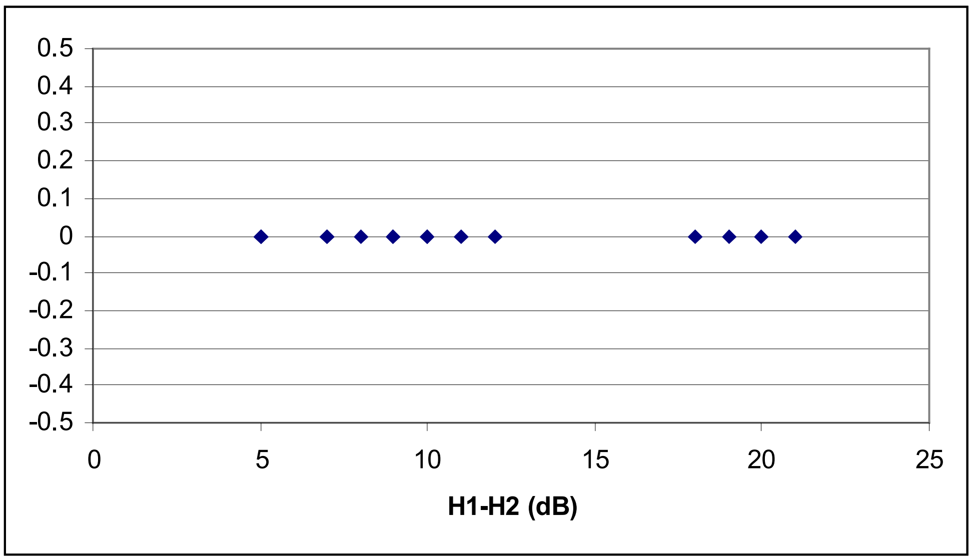


Figure 82: Graph of the measured H1-H2 values of the stimuli from Experiment 1. (A second "dimension" was added to the graph along the y-axis to parallel Figure 81.)

From Figure 82, we can see that there are a few gaps between stimuli clustered along H1-H2. There is a gap between the stimuli with H1-H2 values of 5 dB and the stimulus with an H1-A1 value of 7 dB. In addition, there is a large gap between the stimulus with an H1-H2 value of 12 dB and the stimulus with an H1-H2 value of 18 dB.

Both of these gaps correspond directly to the Gujarati listeners' division along H1-H2 in Figure 81.

4.4.2.2.3 Differences among individual listeners

Table 13, 14, and 15 are of the individual subject weights, as well as the average weight, for English, Spanish, and Gujarati listeners, respectively.

English listeners		
Listeners	**Dimension 1 (H1-H2)**	**Dimension 2 (CPP)**
1	0.33	0.26
2	0.52	0.37
3	0.36	0.28
4	0.41	0.39
5	0.44	0.4
6	0.45	0.34
7	0.44	0.35
8	0.33	0.29
9	0.49	0.4
10	0.51	0.31
11	0.35	0.31
12	0.36	0.32
13	0.27	0.22
14	0.39	0.37
15	0.38	0.31
16	0.46	0.44
17	0.55	0.45
18	0.57	0.44
Average	**.42**	**.35**

Table 13: Importance of individual dimensions to English listeners.

Spanish listeners		
Listeners	**Dimension 1 (H1-A1)**	**Dimension 2 (H1-H2)**
1	0.51	0.49
2	0.54	0.51
3	0.36	0.34
4	0.4	0.39
5	0.54	0.51
6	0.5	0.41
7	0.5	0.48
8	0.53	0.45
9	0.41	0.4
10	0.48	0.45
11	0.46	0.44
12	0.5	0.41
13	0.66	0.5
14	0.65	0.54
15	0.55	0.45
16	0.68	0.32
17	0.58	0.42
18	0.61	0.53
Average	**.53**	**.45**

Table 14: Importance of individual dimensions to Spanish listeners.

Gujarati listeners	
Listeners	**Dimension 1 (H1-H2)**
1	0.81
2	0.8
3	0.8
4	0.72
5	0.85
6	0.79
7	0.82
8	0.8
9	0.8
10	0.79
11	0.81
12	0.75
Average	**.80**

Table 15: Importance of individual dimensions to Gujarati listeners.

As with Experiment 1, results showed that English listeners differed in how they weighted each cue. However, once again, all listeners agreed on the relative order of importance of the dimensions (Table 13). For all 18 listeners considered individually, the first dimension of the group solution was always more important than the second dimension of the group solution. That is, each listener weighed dimension one for the group solution as his or her own individual dimension one.

For Spanish listeners, results were also similar to Experiment 1. Again, while listeners differed in how they weighted each dimension, they agreed on the relative order of importance of the dimensions (Table 14). For all 18 listeners considered individually, the first dimension was always more important than the second. That is, each listener weighed dimension one for the group solution as his or her own individual dimension one.

For the Gujarati listeners, we cannot say anything about the relative importance of dimension beyond that fact that they only used one dimension (i.e., it is all or nothing for the Gujarati listeners). Again, individual MDS analyses were run on the Gujarati listeners to see if each listener used H1-H2 as his/her most important dimension. Results showed that the Gujarati listeners were consistent in that all of the listeners, considered individually, had a one-dimensional perceptual space, and this one dimension correlated with H1-H2 for all 12 listeners.

4.4.2.3 Experiment 3 (Talker 3)

4.4.2.3.1 Multidimensional scaling Analyses

Figures 83-85 show the R^2 and stress values for one to five dimensions for English, Spanish, and Gujarati listeners, respectively. In each graph, an arrow is pointing to the elbow.

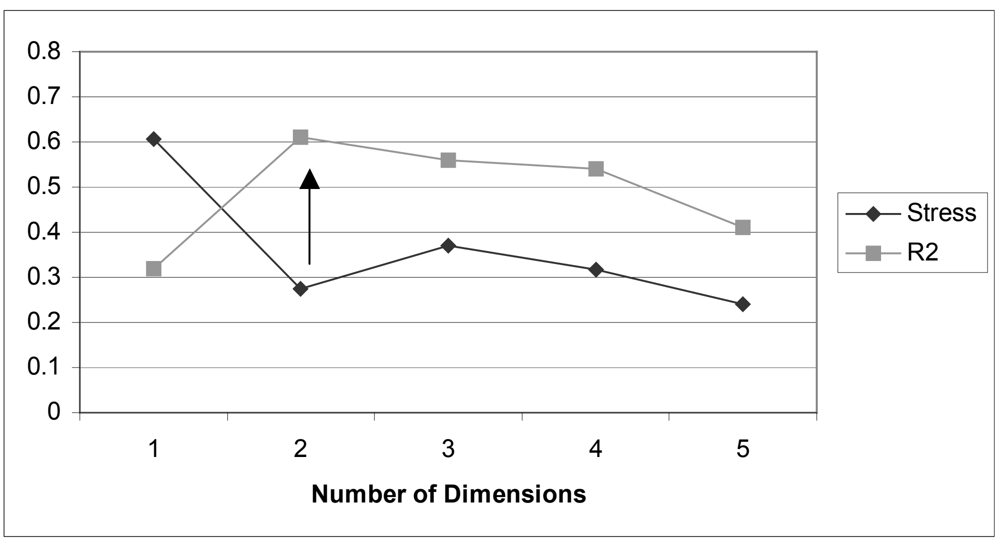


Figure 83: Graph of R^2 and stress for English listeners (Experiment 3).

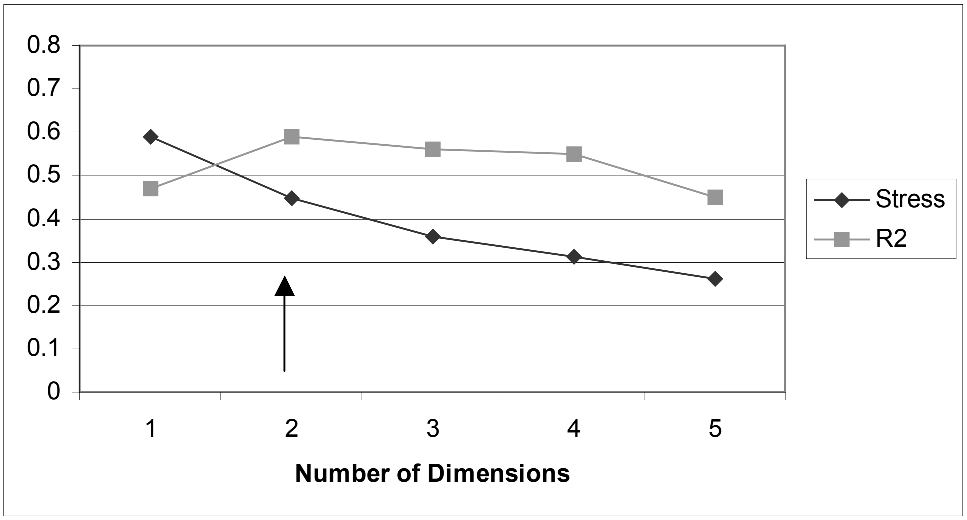


Figure 84: Graph of R^2 and stress for Spanish listeners (Experiment 3).

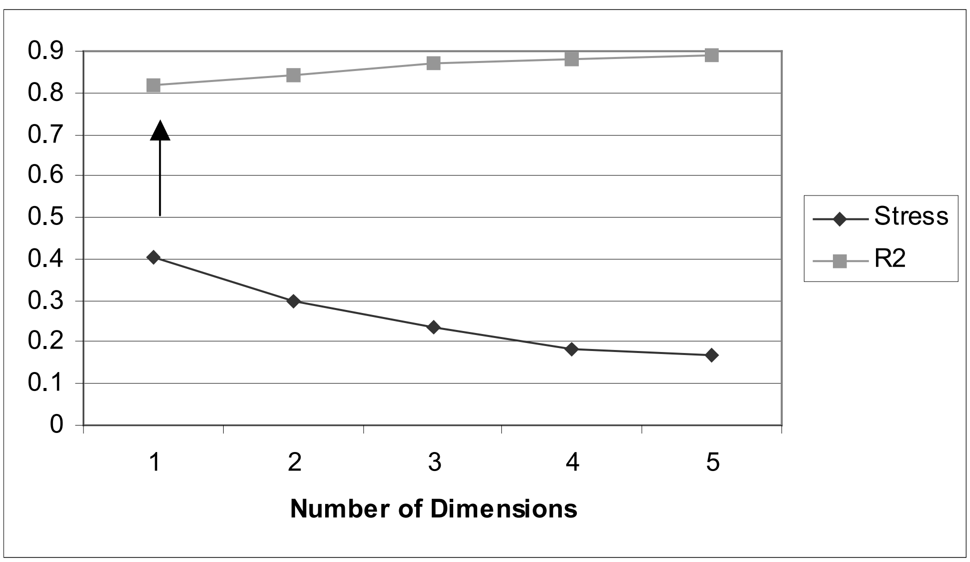


Figure 85: Graph of R^2 and stress for Gujarati listeners (Experiment 3).

For the English listeners, stress and R^2 values indicated a two-dimensional solution, which accounted for 61% of the variance (with a stress value of .27) in the rating data. As with the previous experiment, additional dimensions accounted for less variance.

For the Spanish listeners, stress and R^2 values also indicated a two-dimensional solution, which accounted for 59% of the variance (with a stress value of .45) in the rating data. Again, additional dimensions accounted for less variance.

For the Gujarati listeners, stress and R^2 values indicated a one-dimensional solution, which accounted for 82% variance (with a stress value of .40). After the first dimension, the amount of variance accounted for increased only gradually.

4.4.2.3.2 Group Perceptual Spaces

Figures 86, 88, and 90 represent the perceptual space for Experiment 3 for English, Spanish and Gujarati, respectively. The dimensions were interpreted using the same procedure explained in Experiment 1 (section 4.4.2.1.2).

4.4.2.3.2.1 Perceptual space for English listeners (Experiment 3)

The English listeners' perceptual space is presented in Figure 86, with dimension one graphed on the x-axis, and dimension two on the y-axis.

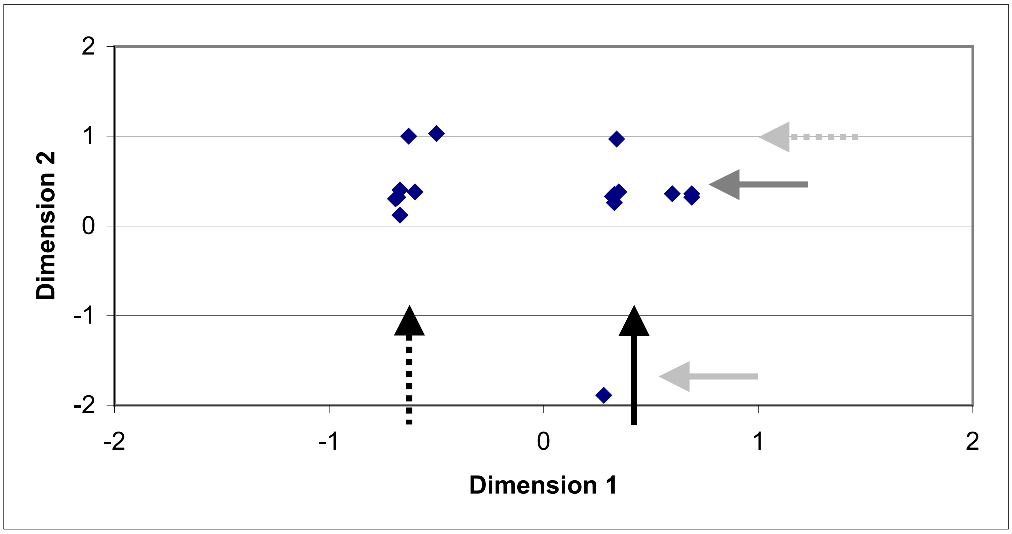


Figure 86: Group perceptual space for English listeners (Experiment 3).

Dimension one represents H1-H2, with the leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-H2 values of approximately 10-19 dB, and the rightmost points (indicated by the solid black arrow) corresponding to stimuli with H1-H2 values less than or equal to 9 dB. There was a significant correlation between dimension one and H1-H2 ($r = .90$, $p < .05$). Dimension two represents CPP with the topmost points (indicated by the dotted gray arrow) representing stimuli with very low CPP values of approximately –1 to 1 dB. The bottom-most point (indicated by the light gray arrow) represents the single stimulus with the greatest CPP value, 9.8 dB. Intermediate points (indicated by the dark gray arrow) represent stimuli with CPP values of approximately 2-6 dB. There are more intermediate points in Figure 86 because there were more stimuli with CPP values between 2-6 dB in Experiment 3 than in the previous two experiments (compare the CPP values of Talker 3 to Talkers 1 and 2 in Figure 48). There was a significant correlation between dimension two and CPP ($r = .78$, $p < .05$).

One could ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 86. Figure 87 is a scatter plot of the measured H1-H2 and CPP values of the stimuli.

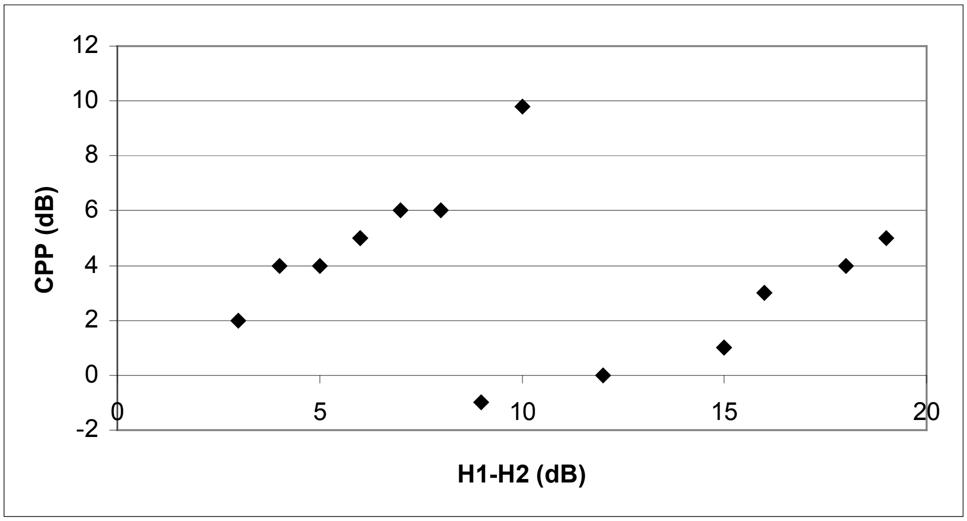


Figure 87: Graph of the measured H1-H2 and CPP values of the stimuli from Experiment 3.

From Figure 87, we can see that, along both dimensions, there are some gaps between the stimuli. However, along the H1-H2 dimension, there are no gaps that correspond directly to the English listeners' division along H1-H2 in Figure 86. But, along the CPP dimension, there is a gap between the stimulus with a CPP value of 9.8 dB and the stimulus with a CPP value of 6 dB. This gap corresponds directly to the English listeners' division along CPP in Figure 86.

4.4.2.3.2.2 Perceptual space for Spanish listeners (Experiment 3)

The Spanish listeners' perceptual space is presented in Figure 88, with dimension one graphed on the x-axis, and dimension two on the y-axis.

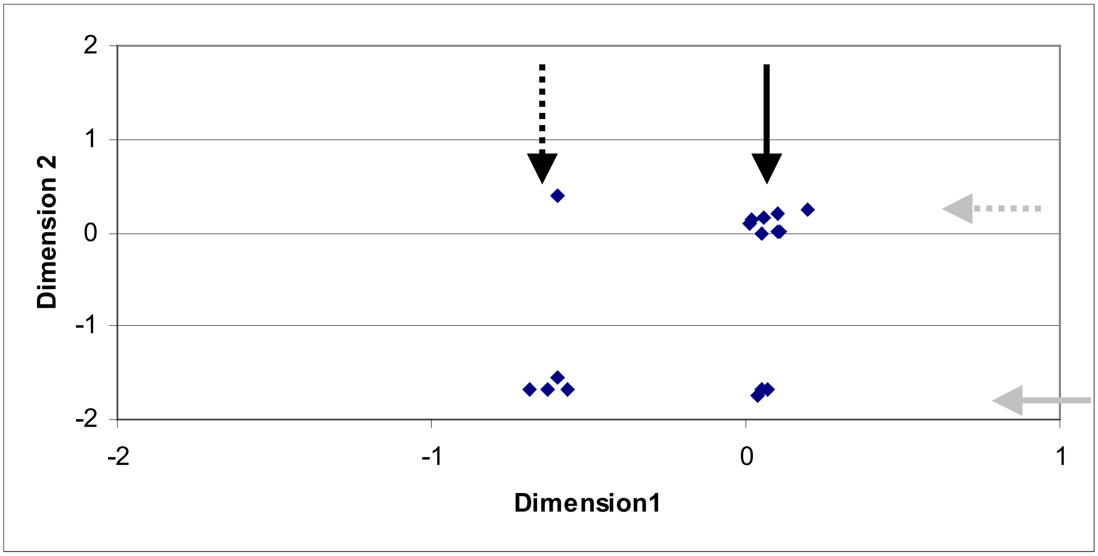


Figure 88: Group perceptual space for Spanish listeners (Experiment 3).

The Spanish listeners organized the perceptual space along two dimensions and divided it into approximately four categories. Dimension one represents H1-A1 with the five leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-A1 values less than or equal to –3 dB. The 11 rightmost points (indicated by the solid black arrow) correspond to stimuli with H1-A1 values greater than or equal to 0 dB. There was a significant correlation between dimension one and H1-A1 ($r = .76$, $p < .05$). Dimension two represents H1-H2, with the nine topmost points (indicated by the dotted gray arrow) representing stimuli with H1-H2 values of approximately 10-18 dB, and the bottom-most points (indicated by the gray arrow) corresponding to H1-H2 values less than or equal to 9 dB. There was a significant correlation between dimension two and H1-H2 ($r = .80$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 88. Figure 89 is a scatter plot of the measured H1-A1 and H1-H2 values of the stimuli.

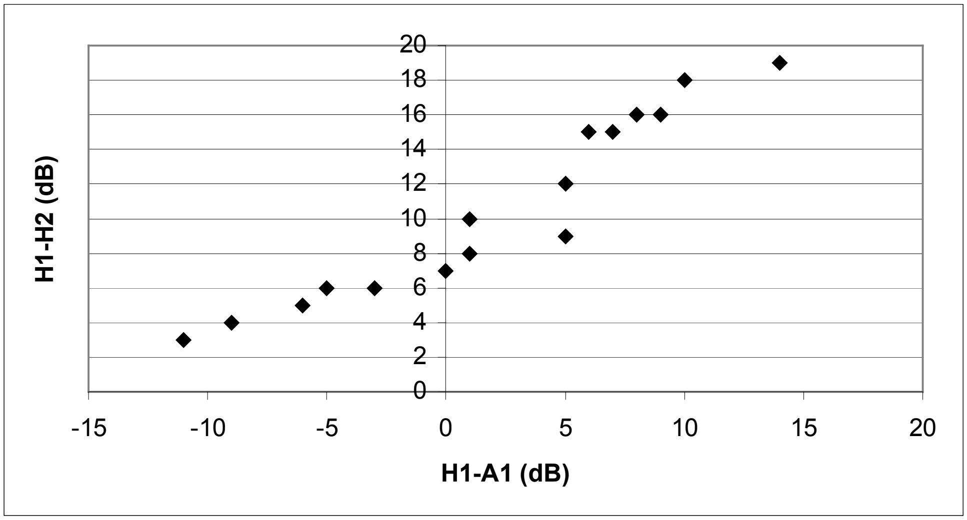


Figure 89: Graph of the measured H1-A1 and H1-H2 values of the stimuli from Experiment 3.

From Figure 89, we can see that, along both dimensions, there are a few gaps between the stimuli. Along the H1-A1 dimension, there is a gap between the stimulus with an H1-A1 value of -3 dB and the stimulus with an H1-A1 value of 0 dB. This corresponds directly to the Spanish listeners' division along H1-A1 in Figure 88. (There is a much larger gap between the stimulus with an H1-A1 value of 1 dB and the stimulus with an H1-A1 value of 5 dB, but the Spanish listeners did not use this gap to make their judgments.) Along the H1-H2 dimension, there are no gaps that correspond directly to the Spanish listeners' division.

4.4.2.3.2.3 Perceptual space for the Gujarati listeners (Experiment 3)

For the Gujarati results (Figure 90) there is only one dimension, which is plotted along the x-axis. However, to make the perceptual space easier to represent graphically, a second “dimension” was added to the graph along the y-axis. Note how the values on dimension 2 are all zero.

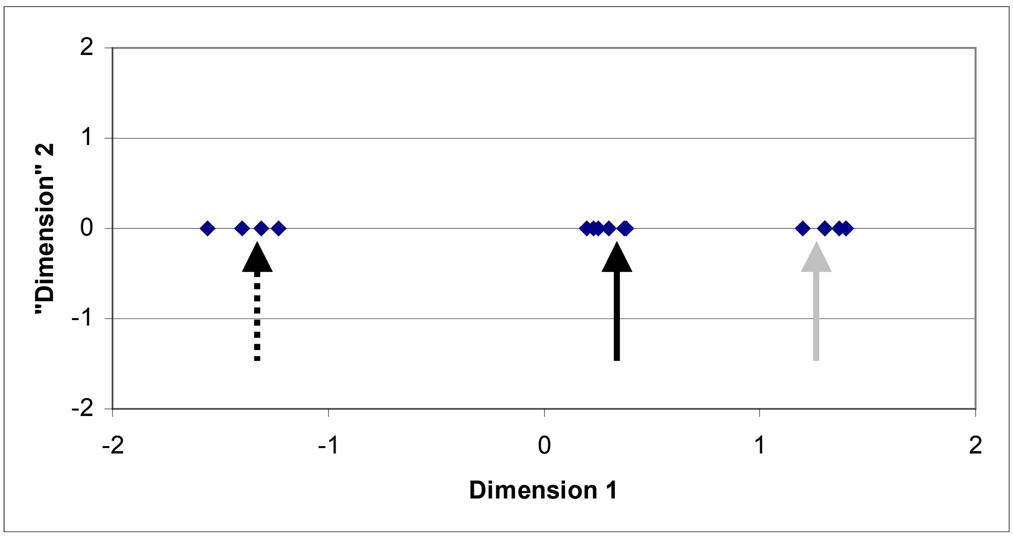


Figure 90: Group perceptual space for Gujarati listeners (Experiment 3).

Again, as we saw with both Experiment 1 and 2, the stimuli clustered into three groups, distinguished by dimension one, which corresponds H1-H2. The four leftmost points (indicated by the dotted black arrow) correspond to stimuli with H1-H2 values greater than or equal to 15 dB, the middle set of points (indicated by the solid black arrow) correspond to stimuli with H1-H2 values between six and 12 dB, and the rightmost point (indicated by the gray arrow) correspond to stimuli with H1-H2 values less than or equal to 5 dB. There was a significant correlation between dimension one and H1-H2 ($r = .96$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 90. Figure 91 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 Figure 90.

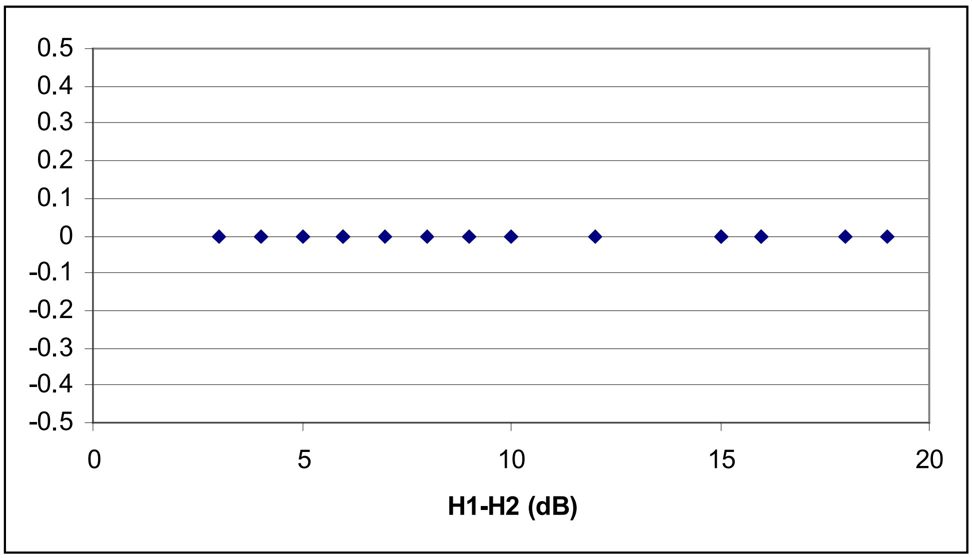


Figure 91: Graph of the measured H1-H2 values of the stimuli from Experiment 3. (A second "dimension" was added to the graph along the y-axis to parallel Figure 90.)

From Figure 91, we can see that there are a few gaps between stimuli clustered along H1-H2. There is a gap between the stimulus with an H1-H2 value of 12 dB and the stimulus with an H1-H2 value of 15 dB. This gap corresponds directly to the Gujarati listeners' division along H1-H2 in Figure 91. However, there are no gaps in the stimuli that correspond to any of the other divisions that the Gujaratis made.

4.4.2.3.3 Differences among Individual Listeners

Tables 16, 17, and 18 are of the individual subject weights, as well as the average weight, for English, Spanish, and Gujarati listeners, respectively.

English listeners		
Listeners	**Dimension 1 (H1-H2)**	**Dimension 2 (CPP)**
1	0.4	0.38
2	0.37	0.29
3	0.52	0.46
4	0.51	0.46
5	0.43	0.26
6	0.53	0.39
7	0.51	0.49
8	0.5	0.47
9	0.54	0.53
10	0.22	0.15
11	0.38	0.32
12	0.47	0.42
13	0.57	0.36
14	0.49	0.39
15	0.51	0.49
16	0.55	0.32
17	0.67	0.21
18	0.57	0.44
Average	**.49**	**.38**

Table 16: Importance of individual dimensions to English listeners.

Spanish listeners		
Listeners	**Dimension 1 (H1-A1)**	**Dimension 2 (H1-H2)**
1	0.50	0.49
2	0.54	0.51
3	0.36	0.34
4	0.4	0.39
5	0.54	0.51
6	0.5	0.41
7	0.49	0.48
8	0.53	0.45
9	0.41	0.4
10	0.46	0.45
11	0.46	0.44
12	0.5	0.41
13	0.66	0.5
14	0.51	0.48
15	0.65	0.6
16	0.62	0.59
17	0.61	0.56
18	0.45	0.43
Average	**.51**	**.47**

Table 17: Importance of individual dimensions to Spanish listeners.

Gujarati listeners	
Listeners	**Dimension 1 (H1-H2)**
1	0.84
2	0.86
3	0.76
4	0.71
5	0.86
6	0.81
7	0.83
8	0.8
9	0.8
10	0.79
11	0.84
12	0.72
Average	**.80**

Table 18: Importance of individual dimensions to Gujarati listeners.

As with Experiments 1 and 2, for English listeners, results show that listeners differed in how they weighted each cue. However, all listeners agreed on the relative order of importance of the dimensions (Table 16). For all 18 listeners considered individually, the first dimension was always more important than the second. That is, each listener weighed dimension one for the group solution as his or her own individual dimension one.

For Spanish listeners, results show that listeners differed in how they weighted each cue. However, all listeners agreed on the relative order of importance of the dimensions (Table 17). For all 18 listeners considered individually, the first dimension was always more important than the second. That is, each listener weighed dimension one for the group solution as his or her own individual dimension one.

For the Gujarati listeners, we cannot say anything about the relative importance of dimension beyond that fact that they only used one dimension (i.e., it is all or nothing for the Gujarati listeners). Again, individual MDS analyses were run on the Gujarati listeners to see if each listener used H1-H2 as his/her most important dimension. Results showed that the Gujarati listeners were consistent in that all of the listeners had a one-dimensional perceptual space, and this one dimension correlated with H1-H2 for all 12 listeners.

4.5 Summary

The goals of this chapter were to answer the following questions:

1) Does linguistic experience affect one's perception of phonation?
2) What acoustic properties correlate with listeners' perception?

It was hypothesized that the results would be similar to the experiment presented in the previous chapter. More specifically, two outcomes were hypothesized. First, Gujarati listeners would be more uniform in their judgments than English and Spanish listeners (that is to say, that Gujaratis would have greater within-listener and cross-listener agreement.) This prediction was partially refuted. Gujarati listeners did have greater within-listener agreement than either English or Spanish listeners. However, English and Spanish listeners were consistent in their judgments across listeners, in that each listener behaved the same way as the entire group of listeners. That is, each listener weighed dimension one for the group solution as his/her own individual dimension one; none of the English or Spanish listeners weighed dimension two over dimension one. The second hypothesis was that all three listener groups would use H1-H2, in addition to H1-A1 for the Spanish listeners. This hypothesis was confirmed by the three experiments presented in this chapter. However, English listeners also use CPP, which was surprising because they did not base their judgments on this measure in the free-sort task, where it was the best predictor in the discriminant analysis.

More specifically, in all sets of three similarity ratings, English listeners based their judgments on H1-H2 and CPP, in that order. Across the three similarity-rating tasks, results were very similar. For example, when using H1-H2, English listeners always mapped the stimuli into roughly two clusters, one with an H1-H2 less than or equal to 9 dB and the other with an H1-H2 value greater than or equal to 9 dB. However, in terms of CPP, judgments were not always as consistent. In Experiments 1 and 3, the stimuli were clustered into roughly three groups, though the CPP values of the groups were different across the experiments. In Experiment 1, the CPP of the three clusters corresponded to 13-15 dB, 5-12 dB, and less than or equal to 5 dB. In Experiment 2, the stimuli were clustered into only two groups, one corresponding to stimuli with CPP values greater than or equal to 8 dB, and the other to stimuli with CPP values less than or equal to 6 dB. In Experiment 3, the stimuli were clustered into three groups with CPP values of greater than 9 dB, 2-6 dB, and 1 dB or less. One reason that the stimuli in Experiment 2 were clustered into only two groups was that there were not many stimuli with intermediate CPP values in this experiment.

In all three similarity-rating tasks, Spanish listeners based their judgments on H1-A1 and H1-H2. In all three experiments, Spanish listeners favored H1-A1 over H1-H2, and, in all three experiments, Spanish listeners used H1-A1 and H1-H2 in similar ways. In Experiment 1, the stimuli clustered into three sections, corresponding to stimuli with

H1-A1 values of –5 to –9 dB, 0 to 8 dB, and 17-21 dB. In Experiments 2 and 3, the stimuli clustered into two groups, one with H1-A1 values less than or equal to 0 dB, and the other group with H1-A1 values greater than 0 dB. In addition, in all three similarity-rating tasks, Spanish listeners used H1-H2 in a similar way. When basing their judgments on H1-H2, Spanish listeners mapped the stimuli in a way similar to the English listeners (that is, the stimuli were clustered into roughly two groups, one with an H1-H2 of less than or equal to 9 dB and the other with an H1-H2 value greater than or equal to 9 dB). With respect to H1-H2, Spanish listeners behaved no differently than English listeners, despite the allophonic breathiness in English.

The Gujarati listeners only used H1-H2, and did so in a way different from either English or Spanish listeners. In all three experiments, the Gujarati listeners' perceptual mapping of the stimuli was clustered into three sections, with very similar H1-H2 values across experiments. In Experiment 1, the three clusters corresponded to stimuli with H1-H2 values of 2-3 dB, 4-12 dB, and 15 dB; in Experiment 2, the three clusters corresponded to stimuli with H1-H2 values of 5 dB, 7-12 dB, and 18-21 dB; in Experiment 3, the three clusters corresponded to stimuli less than or equal to 5 dB, 6-12 dB, and greater than or equal to 15 dB.

Even though Gujarati only has two phonation categories, the Gujaratis consistently mapped the stimuli into three clusters. Perhaps, there is a ceiling for breathy phonation in Gujarati. Anything with H1-H2 values in the range of approximately 1-5 dB are considered "modal", stimuli with H1-H2 values of approximately 6-12 dB are "breathy", and stimuli with H1-H2 values greater than 15 dB are "other". This division corresponds to the breathy versus modal distinction in Gujarati. In the sample measured for this study, the H1-H2 value for modal vowels ranged from –1-4 dB, while the H1-H2 value for breathy vowels ranged from 7-12 dB. There were no stimuli with H1-H2 values higher than 12 dB in the sample measured for this study.

Furthermore, the perceptual mapping of the stimuli by the Gujaratis did not correspond to the breathy versus modal distinction in Mazatec. First, the Gujaratis always mapped the stimuli into three clusters, despite the fact that there are only two categories in Mazatec. Second, Gujaratis sometimes judged stimuli with two different phonation categories to be perceptually similar. For example, Talker 2 (see Figure 50) produced categorically modal vowels with high H1-H2 values (around 8 dB), but the Gujaratis judged these stimuli to be perceptually more similar to the Mazatec breathy vowels. The perceptual mapping of the stimuli reflected the breathy versus modal distinction in Gujarati, to some extent (again, in the sample measured for this study, the H1-H2 value for modal vowels ranged from –1-4 dB, while the H1-H2 value for breathy vowels ranged from 7-12 dB).

One question that was addressed in this chapter was whether the acoustic properties of the stimuli naturally lent themselves to the division that the listeners made. We saw that, at times, this was the case. For example, across the three experiments, when relying on H1-A1, the Spanish listeners' division of the stimuli did correspond to a natural division in the stimulus set. However, when using H1-H2, the Spanish listeners' judgments did not correspond to any natural division in the stimulus set. Across the three experiments, the Spanish listeners always divided the stimuli into two clusters (one with an H1-H2 value of 9 db or greater, and the other with an H1-H2 value less than 9 dB), but the stimuli never lent themselves to this division. For English listeners, results were similar. Along H1-H2, the English listeners' division of the stimuli never corresponded to any gaps in the stimulus set. However, along CPP, some of the English listeners' divisions did correspond to gaps in the stimuli. For Gujarati listeners, results were mixed; only at times did their division of the stimuli correspond to gaps in the stimulus set.

Finally, we saw that listeners did not rely on the measures that the discriminant analysis predicted to be the most successful. Recall that the results of the discriminant analysis showed that H1-A2 accounted for most of the variance (53%) in the stimulus set, followed by H1-A1 (20%) and then H1-H2 (14%). None of the listener groups studied in this experiment based their judgments on H1-A2, despite the fact that it accounted for most of the variance in the stimulus set. However, the Spanish listeners did rely on the measures that accounted for the second and third most variance in the stimulus set (H1-A1 and H1-H2, in that order).

Chapter 5: Perception of pathologically-disordered voices

5.1 Introduction

In this chapter, the following questions will be addressed:

1) Are listeners who are sensitive to linguistically-relevant phonation differences more sensitive to pathologically-disordered phonation?
2) What acoustic properties correlate with listeners' perception of pathologically-disordered phonation?

As previously mentioned, there has not been any work done on the perception of pathologically-disordered phonations using listeners of languages with phonation contrasts. However, we know from previous research that at least some pathologically-disordered phonations resemble linguistically-relevant ones (e.g. Ladefoged 1983; Epstein 1999). Therefore, one might hypothesize that Gujaratis will be more consistent in their judgments of pathologically-disordered phonations based on the breathy phonation category in their own language, and that they will use H1-H2 to make their judgments. Based on the results from the three experiments in Chapter 4, it is hypothesized that English listeners will judge pathologically-disordered phonations using H1-H2 and CPP, while Spanish listeners will use H1-A1 and H1-H2. However, it is believed that English and Spanish listeners will not be consistent in their judgments of the voices.

The organization of this chapter is as follows: section 5.2 will explain the methods used in this study, including a description of the listeners (section 5.2.1) and the stimuli (section 5.2.2.), section 5.3 will explain the general procedure for the similarity-rating task, section 5.4 will present of the results, and section 5.5 will summarize the results of the chapter.

5.2 Methods

5.2.1 Listeners

The same listeners who participated in the experiments presented in Chapter 3 and Chapter 4 participated in this task.

5.2.2 Stimuli

5.2.2.1 Pathologically-disordered stimuli

The voices of 16 female native speakers of American English with a variety of voice disorders were selected from a library of samples provided by the Bureau of Glottal Affairs, Division of Head/Neck Surgery, UCLA School of Medicine. Each talker was recorded sustaining the vowel [a] for as long as possible. (Each talker contributed one stimulus.) Of the 16 voices, an attempt was made to select 14 that had similar spectral properties to the Mazatec stimuli (used in the experiment in Chapter 3) so that a direct comparison of the experiments could be made. For example, the H1-H2 values of the Mazatec stimuli range from 3-22 dB (across the three talkers). Fourteen (of the sixteen) pathologically-disordered stimuli with H1-H2 values in the range of 3-22 dB were selected for the experiment. The other two pathologically-disordered stimuli had more extreme values.

5.2.2.2 Manipulation of the stimuli

Each vowel was normalized to 250 ms by cutting out individual pulses using the same procedure explained in section 3.2.2.2. F0 was normalized to a falling f0 of 210-

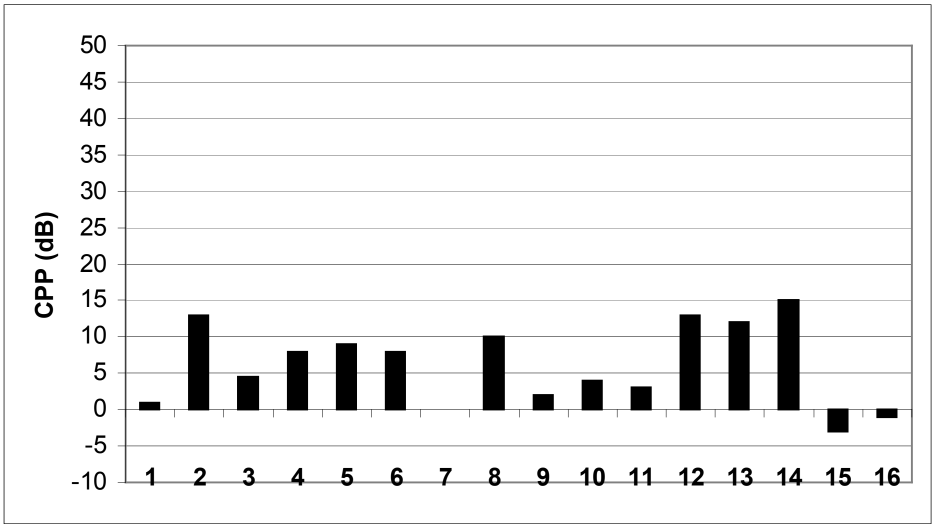
200 Hz using the PSOLA (pitch-synchronous overlap and add) function of Praat software.

5.2.2.3 Phonetic measurements of the stimuli

Acoustic measures were made on all of the stimuli to determine which dimension(s) might correlate with listeners' perception. Only acoustic measures were made since the stimuli were only available as audio signals. CPP, H1-H2, H1-A1, H1-A2, H1-A3, the average of H1-H2 compared to A1, A2-A3, and H2-H4 were measured for all of the stimuli following the same procedure explained in section 3.3.2.3.

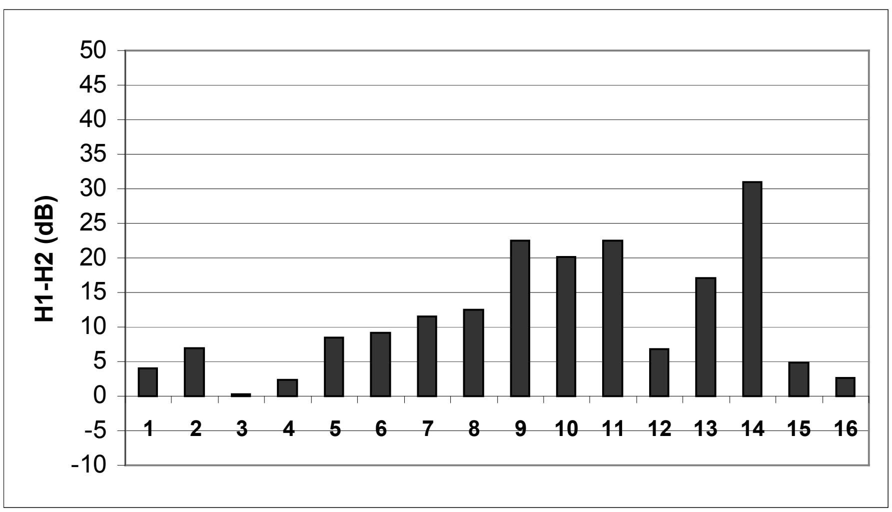
5.2.2.3.1 Results

The results of the acoustic measures for the stimulus from each talker are presented in Figures 92-98. In Figure 92, the height of the cepstral peak is plotted along the y-axis in dB. In Figures 93-98, the difference between the amplitudes of the harmonics (H1-H2, H1-A3, H1-A1, H1-A2, the average of H1-H2 compared to A1, A2-A3, and H2-H4) is given on the y-axis in dB. Samples are given on the x-axis. In each figure, an arrow is pointing in the direction of increased breathiness.



increased
breathines

Figure 92: Graph of individual CPP values for 16 pathologically-disordered voices.



increased
breathiness

Figure 93: Graph of individual H1-H2 values for 16 pathologically-disordered voices.

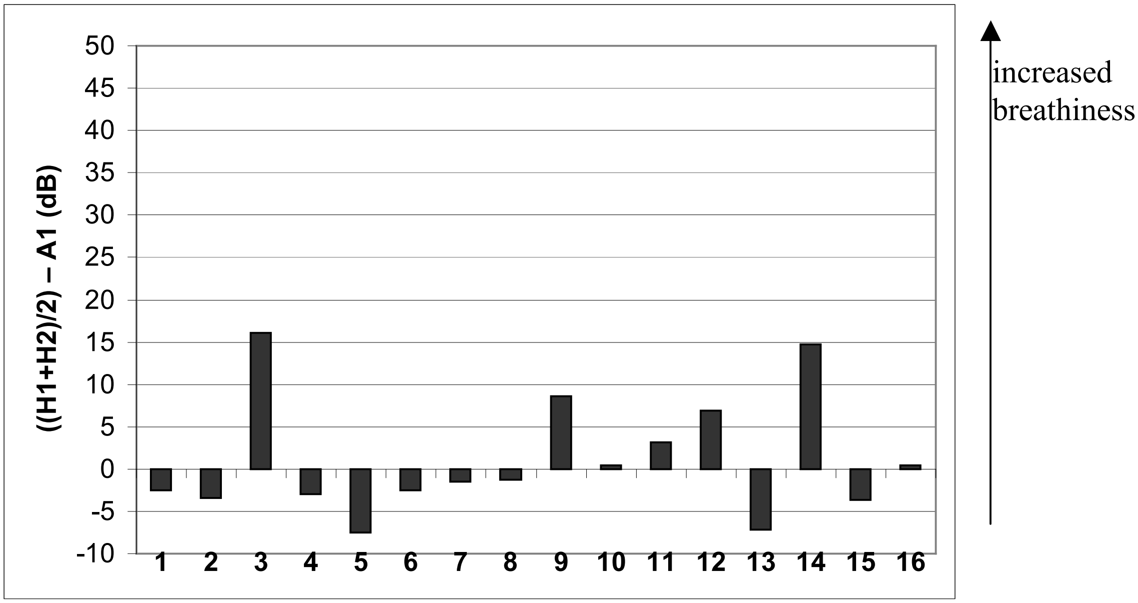


Figure 94: Graph of individual ((H1+H2)/2) – A1 values for 16 pathologically-disordered voices.

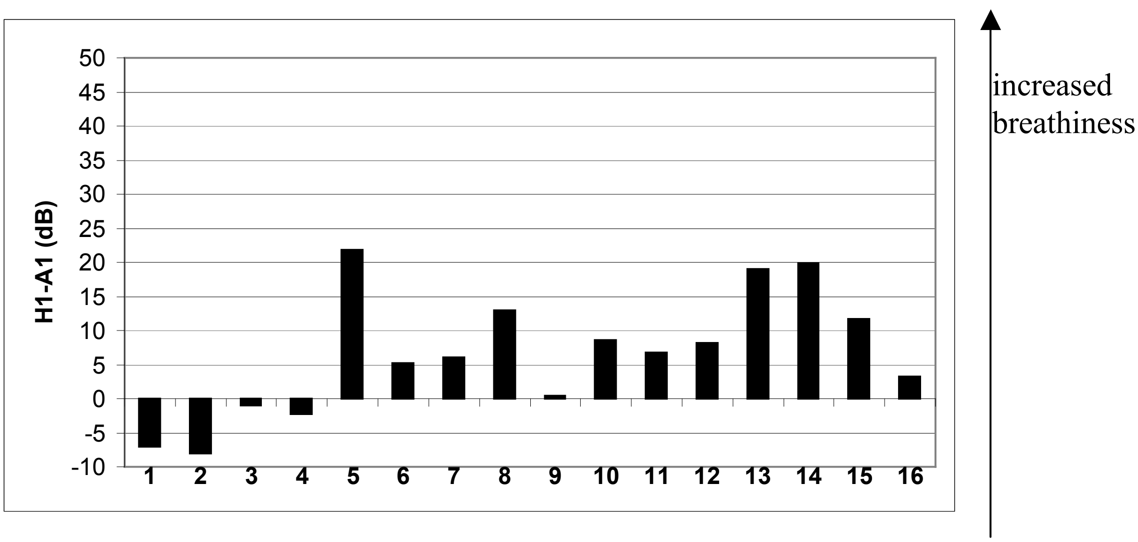


Figure 95: Graph of individual H1-A1 values for 16 pathologically-disordered voices.

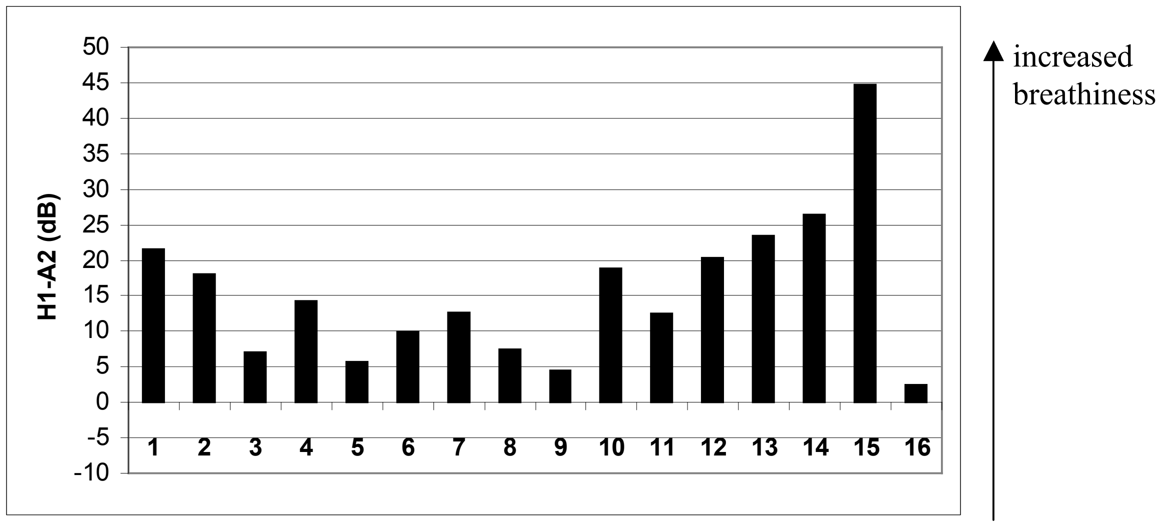


Figure 96: Graph of individual H1-A2 values for 16 pathologically-disordered voices.

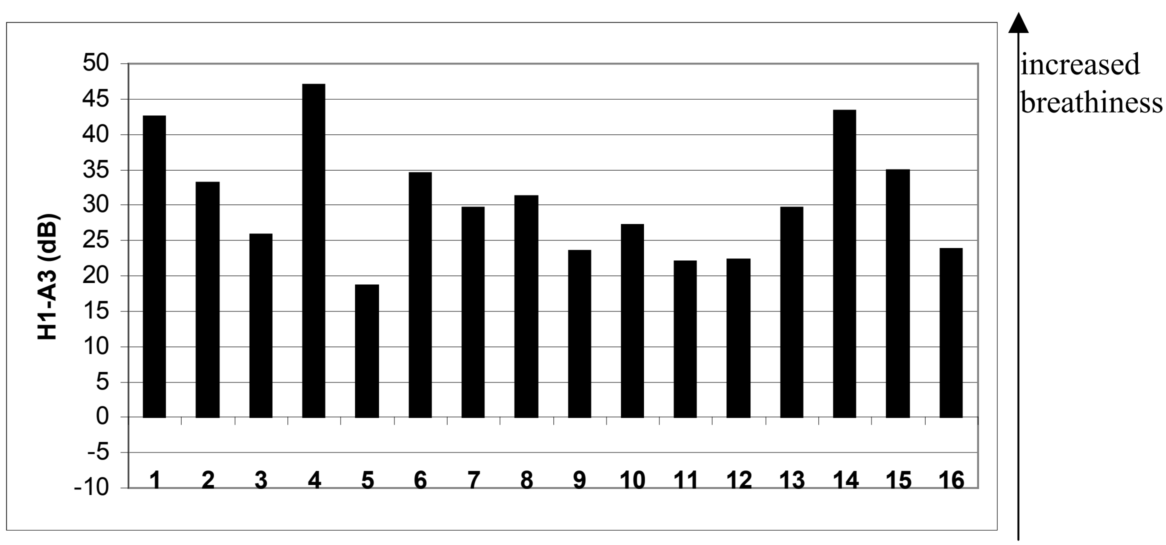


Figure 97: Graph of individual H1-A3 values for 16 pathologically-disordered voices.

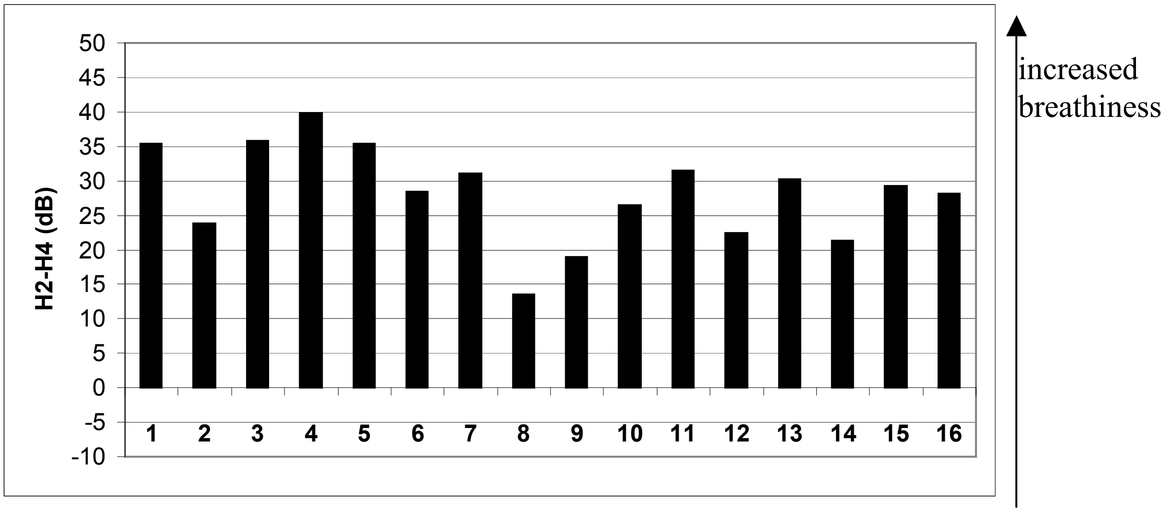


Figure 98: Graph of individual H2-H4 values for 16 pathologically-disordered voices.

We can not say anything about the success of the various measures on the stimuli presented above because 1) there is only one stimulus per talker and 2) the stimuli can not be categorized into breathy or modal groups (i.e. they are all pathologically-disordered). However, we can see that, along certain measures, some of the stimuli have values that are typically associated with modal or breathy phonation. For example, Talker 3 produced stimuli with low H1-H2 values (e.g., 1 db, 2 dB, etc.), which are typically associated with modal phonation, and other stimuli with high H1-H2 value (e.g., 10 dB, 20 dB, etc.), typically associated with breathy phonation.

5.3 Procedure

Each listener participated in one similarity-rating task, which was randomly ordered with respect to the tasks presented in Chapters 3 and 4. The same procedure described in section 4.3 was used here. Listeners began with a practice session composed of eight pathologically-disordered stimuli not used in the experiment. Again, the directions were given in the listeners' native language following the same procedure explained in Chapter 3. As with the experiments presented in Chapter 4, 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”

Responses were analyzed using MDS, and the MDS results were analyzed following the same procedures outlined in section 4.3. Again, within-listener agreement and separate INDSCAL solution were determined for each group of listeners.

5.4 Results

5.4.1 Within-listener agreement

The results of the average within-listener reliability for each language are presented in Table 19. The range of values is also presented. The percent of responses for a single listener’s rating for the AB and BA order of a pair of voices that differed by a value of 40 or more is labeled “% ratings > +/- 40 scale value” and the correlation between the first and the second rating of each pair of stimulus is labeled “r for 1st and 2nd ratings”.

Pathologically-disordered voice task				
	% ratings > +/- 40 scale value		r for 1st and 2nd ratings	
Listeners	**mean**	**range**	**mean**	**range**
English	20.4%	5.3-25.4%	.63	.47-.75
Spanish	22.1%	2.4-26.0%	.58	.43-.70
Gujarati	5.0%	3.2-10.0%	.89	.67-.94

Table 19: Within-listener reliability for English, Spanish, and Gujarati listeners.

Overall, within-listener agreement was good. On average, English listeners deviated by more than a scale value of 40 on only 20.4% of the pairs. In addition, the rating and re-rating of pairs (that is, AB and BA) were correlated for the English listeners ($r = .63$).

On average, Spanish listeners deviated by more than a scale value of 40 on only 22.1% of the pairs. Once again, we see that Spanish listeners' results were very similar to the English listeners', despite the allophonic breathiness in English. In addition, the rating and re-rating of pairs (that is, AB and BA) were correlated for the Spanish listeners ($r = .58$).

As expected, Gujarati listeners were more consistent in their judgments than either English or Spanish speakers. The Gujarati listeners, on average, deviated by more than a scale value of 40 on only 5% of the pairs. In addition, the rating and re-rating of pairs (that is, AB and BA) were well correlated for the Gujarati listeners ($r = .89$).

5.4.2 Multidimensional scaling solution

As with the previous experiment, listeners' responses were analyzed using the INDSCAL model of MDS. Separate solutions in one to five dimensions were found for each listener group. R^2 and stress values were used to find the location of the elbow. Figures 99-101 show the R^2 and stress values for a one to five dimension solution for English, Spanish, and Gujarati listeners, respectively. The number of dimensions is plotted on the x-axis. The R^2 and stress values are plotted on the y-axis in an unspecified value. In each graph, an arrow is pointing to the elbow.

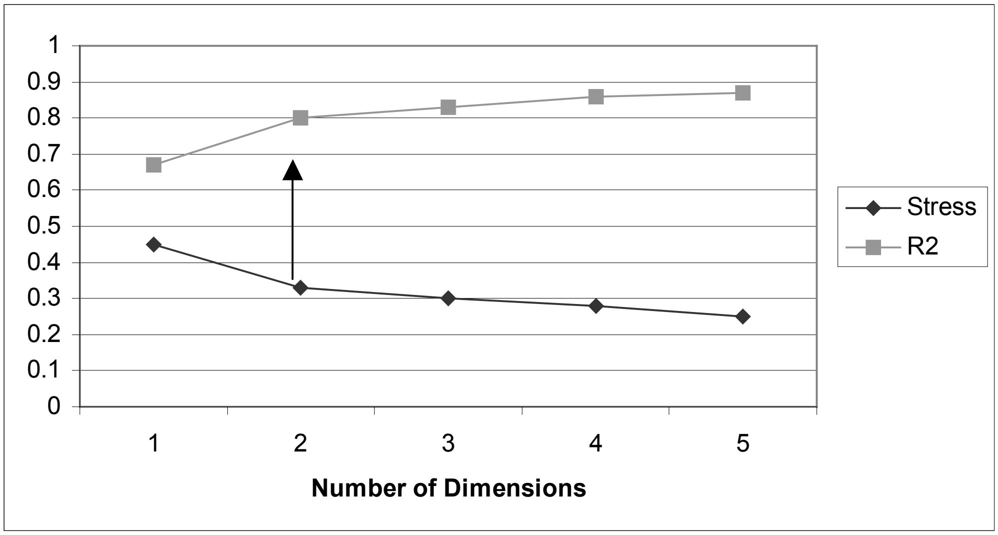


Figure 99: Graph of R^2 and stress for English listeners.

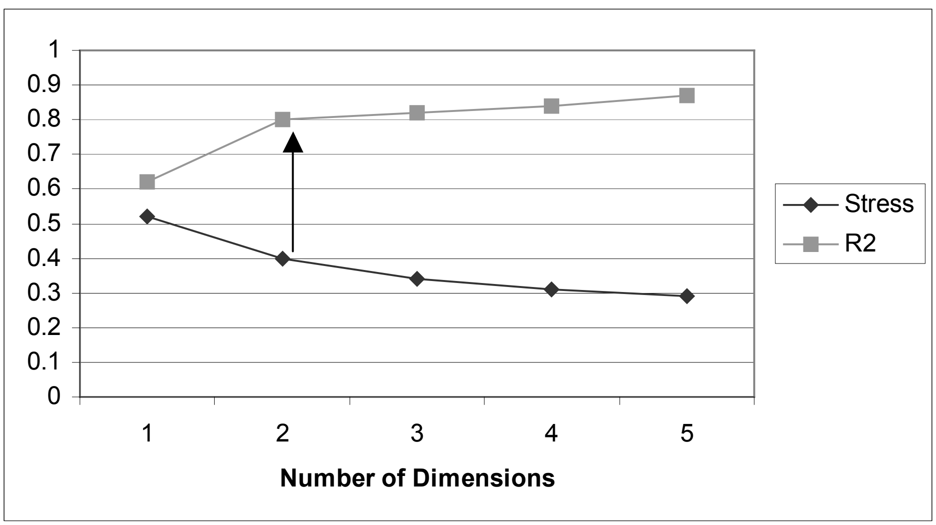


Figure 100: Graph of R^2 and stress for Spanish listeners.

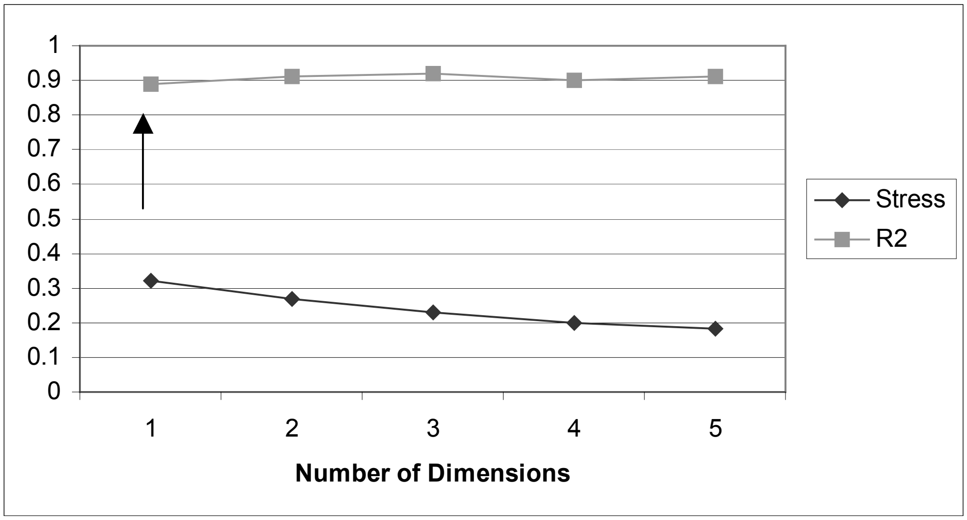


Figure 101: Graph of R^2 and stress for Gujarati listeners.

For the English listeners, stress and R^2 values indicated a two-dimensional solution, which accounted for 80% of the variance (with a stress value of .33) in the data. Most of the variance was accounted for by one dimension (67%, with a stress value of .45). Adding a second dimension accounted for an addition 13% of the variance. After the second dimension, the amount of variance accounted for increased only gradually.

For the Spanish listeners, stress and R^2 values also indicated a two-dimensional solution, which accounted for 80% of the variance (with a stress value of .40) in the data. Most of the variance was accounted for by one dimension (62%, with a stress value of .52). Adding a second dimension accounted for an additional 18% of the variance. After the second dimension, the amount of variance accounted for increased only gradually.

For the Gujarati listeners, stress and R^2 values indicated a one-dimensional solution, which accounted for 89% of the variance in the rating data (with a stress value of .32). After the first dimension, the amount of variance accounted for increased very little.

5.4.2.1 Group perceptual spaces

Figures 102, 104, and 106 represent the perceptual spaces for English, Spanish and Gujarati listeners, respectively. For all three graphs, each point represents one of the 16 stimuli in the experiment. 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.

5.4.2.1.1 Perceptual space for English listeners

The English listeners' perceptual space is presented in Figure 102, with dimension one graphed on the x-axis, and dimension two on the y-axis.

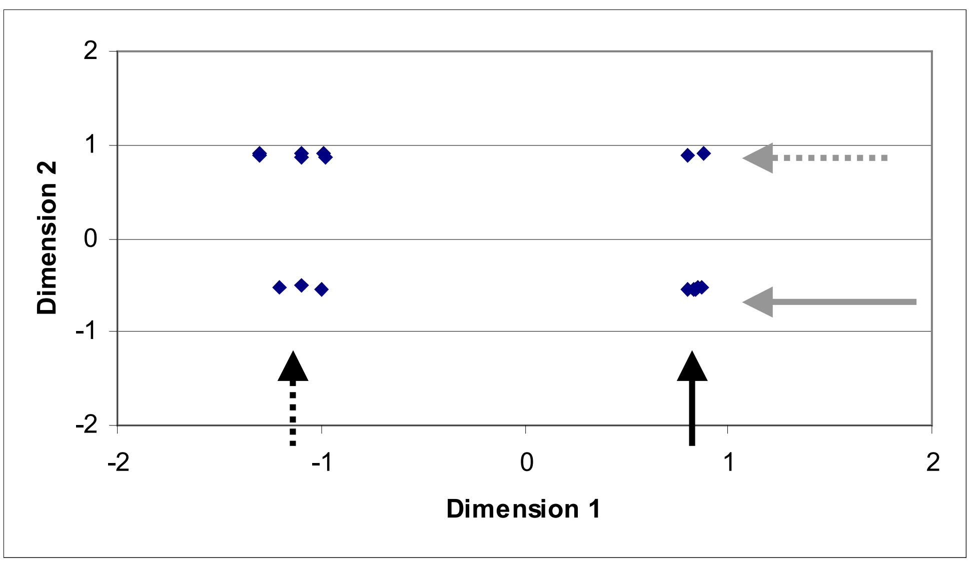


Figure 102: Group perceptual space for English listeners

The English listeners organized the perceptual space along two dimensions and divided it into approximately four categories. Dimension one represents H1-H2, with the leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-H2 values of approximately .3-9.1 dB, and the rightmost points (indicated by the solid black arrow) corresponding to stimuli with H1-H2 values of approximately 11.5-31 dB. There was a significant correlation between dimension one and H1-H2 ($r = .73$, $p < .05$). Dimension two represents CPP with the topmost points (indicated by the dotted gray arrow) corresponding to stimuli with CPP values of approximately –3 to 4.5 dB. The bottom-most points (indicated by the solid gray arrow) represent stimuli with lower CPP values of approximately 8-15 dB. There was a significant correlation between dimension two and CPP ($r = .80$, $p < .05$).

We can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 102. Figure 103 is a scatter plot of the measured H1-H2 and CPP values of the stimuli.

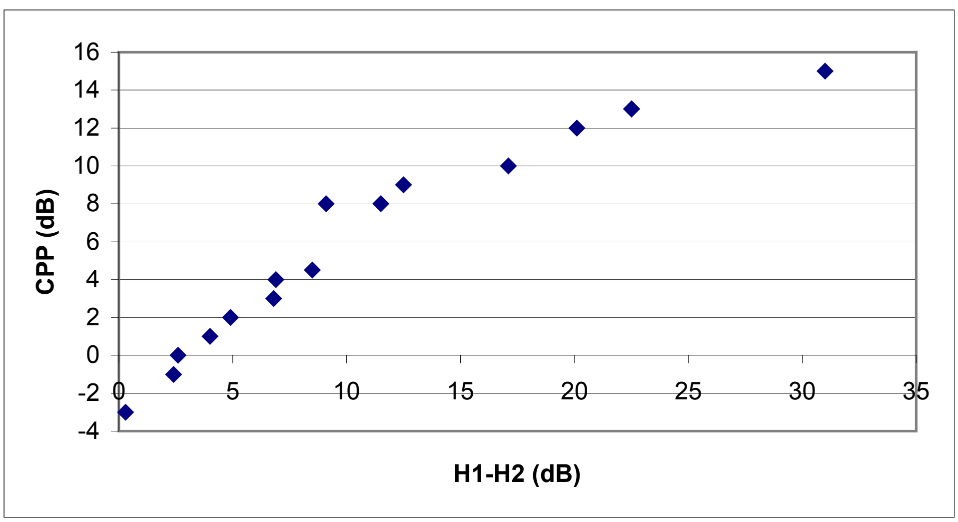


Figure 103: Graph of the measured H1-H2 and CPP values for the pathologically-disordered voice stimuli.

From Figure 103, we can see that, along both dimensions, there are some gaps between the stimuli. Along the H1-H2 dimension, there is a gap between the stimulus with an H1-H2 value of 9.1 dB and the stimulus with an H1-H2 value of 11.5 dB. This corresponds directly to the English listeners' division along H1-H2 in Figure 102. Along the CPP dimension, there is a large gap between the stimulus with a CPP value of 4.5 dB and the stimulus with a CPP value of 8 dB. This corresponds directly to the English listeners' division along CPP in Figure 102.

5.4.2.1.2 Perceptual space for Spanish listeners

The Spanish listeners' perceptual space is presented in Figure 104, with dimension one graphed on the x-axis, and dimension two on the y-axis.

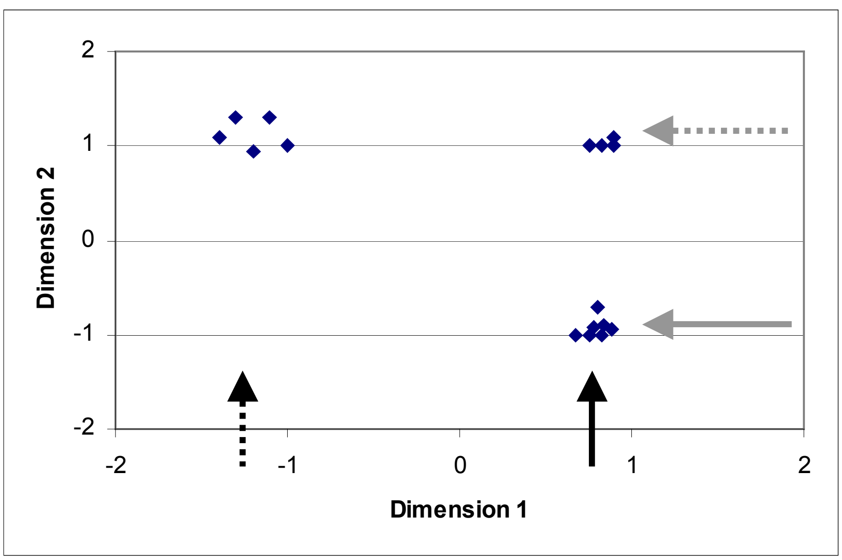


Figure 104: Group perceptual space for Spanish listeners.

Like the English listeners, Spanish listeners had 2 dimensions, but unlike the English listeners, the Spanish listeners had only 3 clusters. Dimension one represents H1-A1. Here the points are cluster into two groups, with the leftmost points (indicated by the dotted black arrow) corresponding to stimuli with H1-A1 values of approximately -8 to .4 dB, and the rightmost points (indicated by the solid black arrow) corresponding to stimuli with H1-A1 values of approximately 3.3-21.8 dB. There was a significant correlation between dimension one and H1-A1 ($r = .82$, $p < .05$). Dimension 2 represents H1-H2. The topmost points (indicated by the dotted gray arrow) represent stimuli with H1-H2 values of approximately .3-9.1 dB, and the bottom-most points (indicated by the gray arrow) corresponding to stimuli with H1-H2 values of approximately 11.5-31 dB. There was a significant correlation between dimension two and H1-H2 ($r = .81$, $p < .05$).

Again, we can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 104. Figure 105 is a scatter plot of the H1-A1 and H1-H2 values of the stimuli.

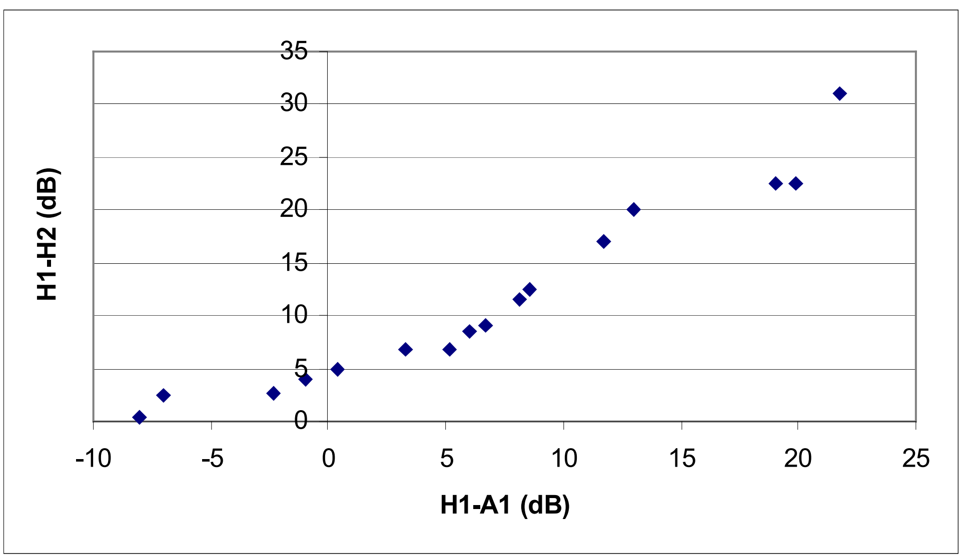


Figure 105: Graph of the H1-A1 and H1-H2 values for the pathologically-disordered voice stimuli.

From Figure 105, we can see that, along both dimensions, there are gaps between the stimuli. Along the H1-H2 dimension, there is a gap between the stimulus with an H1-H2 value of 9.1 dB and the stimulus with a H1-H2 value of 11.5 dB. This corresponds directly to the Spanish listeners' division along H1-H2 in Figure 102. There is also a gap between the stimulus with an H1-A1 value of .4 dB and the stimulus with an H1-A1 value of 3.3 dB. This corresponds directly to the Spanish listeners' division along H1-A1 in Figure 102.

5.4.2.1.3 Perceptual space for the Gujarati listeners

For the Gujarati results (Figure 106) there is only one dimension, which is plotted along the x-axis. However, to make the perceptual space easier to represent graphically, a second "dimension" was added to the graph along the y-axis. Note how the values along dimension 2 are all zero.

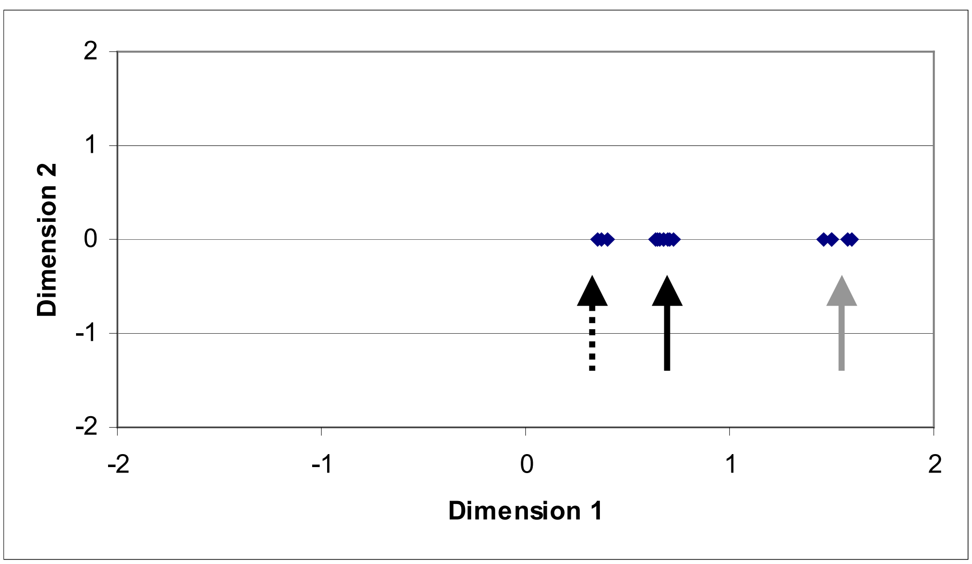


Figure 106: Group perceptual space for Gujarati listeners.

In Figure 106, the stimuli cluster into approximately three groups, distinguished by dimension one, which represents H1-H2. The leftmost points (indicated by the dotted black arrow) correspond to stimuli with H1-H2 values of approximately .3-4 dB, the middle set of points (indicated by the solid black arrow) correspond to stimuli with H1-H2 values between 4.9-12.5 dB, and the rightmost points (indicated by the solid gray arrow) correspond to stimuli with H1-H2 values of 17.1-31 dB. There was a significant correlation between dimension one and H1-H2 ($r = .93$, $p < .05$).

We can ask whether or not the acoustic properties of the stimuli naturally lend themselves to the perceptual mapping presented in Figure 106. Figure 107 is a scatter plot of the H1-H2 values of the stimuli. A second "dimension" was added to the graph along the y-axis to parallel Figure 106.

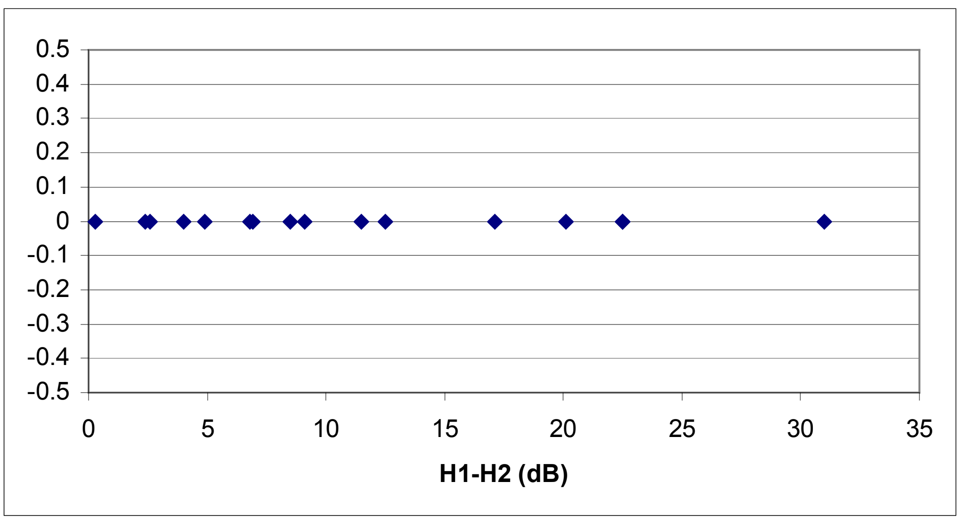


Figure 107: Graph of the H1-H2 values for the pathologically-disordered voice stimuli. (A second "dimension" was added to the graph along the y-axis to parallel Figure 106.)

From Figure 107, we can see that 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 (Figure 106). One exception to this is a large gap between the stimulus with an H1-H2 value of 12.5 dB and the stimulus with an H1-H2 value of 17.1 dB. This corresponds directly to one of the Gujarati listeners' divisions along H1-H2 (Figure 106).

5.4.2.2 Difference among individual listeners

Tables 20, 21, and 22 present the individual subject weights for English, Spanish, and Gujarati listeners, respectively. The average value of the weight for each dimension is given at the end of each chart.

English listeners		
Subject	**Dimension 1 (H1-H2)**	**Dimension 2 (CPP)**
1	0.42	0.56
2	0.63	0.13
3	0.32	0.78
4	0.33	0.54
5	0.1	0.32
6	0.45	0.31
7	0.57	.15
8	0.55	0.47
9	0.67	0.52
10	0.11	0.25
11	0.18	0.26
12	0.72	0.21
13	0.34	0.47
14	0.23	0.34
15	0.55	0.35
16	0.21	0.12
17	0.57	0.44
18	0.45	0.41
Average	**.41**	**.37**

Table 20: Importance of individual dimensions to English listeners.

Spanish listeners		
Subject	**Dimension 1 (H1-A1)**	**Dimension 2 (H1-H2)**
1	0.34	0.48
2	0.23	0.52
3	0.31	0.1
4	0.45	0.21
5	0.56	0.15
6	0.76	0.31
7	0.52	0.53
8	0.32	0.46
9	0.41	0.13
10	0.46	0.56
11	0.57	0.44
12	0.58	0.5
13	0.62	0.43
14	0.32	0.56
15	0.24	0.34
16	0.35	0.49
17	0.61	0.26
18	0.47	0.59
Average	**.45**	**.39**

Table 21: Importance of individual dimensions to Spanish listeners.

Gujarati listeners	
Subject	**Dimension 1 (H1-H2)**
1	0.80
2	0.85
3	0.87
4	0.81
5	0.82
6	0.91
7	0.82
8	0.81
9	0.78
10	0.91
11	0.81
12	0.89
Average	**.84**

Table 22: Importance of individual dimensions to Gujarati listeners.

As Table 20 shows, English listeners differed in how they weighted each cue. Listeners 2, 6, 7, 8, 9, 12, 15, 16, 17, and 18 all weighed dimension one as more important than dimension two, while listeners 1, 3, 4, 5, 10, 11, 13, and 14 did the opposite.

Like the English listeners, Spanish listeners also differed in how they weighted each cue (Table 21). Listeners 3, 4, 5, 6, 9, 11, 12, 13, and 17 all weighed dimension one as more important than dimension two, while listeners 1, 2, 7, 8, 10, 14, 15, 16, and 18 did the opposite. For the Spanish listeners, the subgroup that weighed dimension one as the most important dimension did not appear to have anything in common (e.g. aspects of language background, age, exposure to English, etc.) that would distinguish them from the subgroup that weighed dimension two as the most important dimension.

For the Gujarati listeners, individual MDS analyses were run 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 had a one-dimensional perceptual space, and this one dimension correlated with H1-H2 for all 12 listeners.

5.5 Summary

This chapter concerned itself with the following questions:

1) Are listeners who are sensitive to linguistically-relevant phonation differences more sensitive to pathologically-disordered phonation?
2) What acoustic properties correlate with listeners' perception of pathologically-disordered phonation?

It was hypothesized that English listeners would judge the stimuli using H1-H2 and CPP, while Spanish listeners would use H1-A1 and H1-H2. The results of this experiment confirmed these hypotheses. In addition, it was hypothesized that English and Spanish listeners would not be consistent in their judgments of the voices. This hypothesis was partially confirmed.

English listeners based their judgments on H1-H2 and CPP, and while within-listener agreement was good, listeners were not consistent with one another, in that they differed in which dimension they weighed more heavily. When relying more on H1-H2, English listeners mapped the stimuli into roughly two clusters, one with H1-H2 values of .3-9.1 dB, and the other with H1-H2 values of 11.5-31 dB. When relying more on CPP, English listeners also mapped the stimuli into roughly two clusters, one with CPP values of –3 to 4.5 dB, and the other with CPP values of 8-15 dB. Along both the CPP and the H1-H2 dimensions, the English listeners' divisions between the clusters roughly corresponded to one of the divisions in the stimulus set.

Spanish listeners based their judgments on H1-A1 and H1-H2. As with English listeners, within-listeners agreement was good, but listeners disagreed as to which dimension they weighed more heavily. When relying more on H1-A1, Spanish listeners mapped the stimuli into roughly two clusters, one with H1-A1 values of –8 to .4 dB, and the other with H1-H2 values of 3.3-21.8 dB. When relying more on H1-H2, Spanish listeners again mapped the stimuli into roughly two clusters, one with H1-H2 values of 1-9 dB, and the other with H1-H2 values of 11-30 dB. Along both the H1-A1 and the H1-H2 dimensions, the Spanish listeners' divisions between the clusters roughly corresponded to one of the divisions in the stimulus set.

For the Gujarati listeners, it was hypothesized that their judgments would be more consistent, and that they would base their judgments on H1-H2. These hypotheses were confirmed. The Gujarati listeners used H1-H2. Furthermore, the Gujarati listeners were consistent; within-listener agreement was higher than it was for English and Spanish listeners and all 12 Gujarati listeners had a one-dimensional perceptual space that correlated with H1-H2.

Furthermore, Gujarati listeners used H1-H2 in a different way than either English or Spanish listeners. Gujarati listeners' perceptual mapping of the stimuli was clustered into three sections corresponding to stimuli with H1-H2 values of .3-4 dB, 4.9-12.5 dB, and 17.1-31 dB. (Recall that there was a gap between the stimulus with a measured H1-H2 value of 12.5 dB and the stimulus with a measured H1-H2 value of 17.1 dB, which corresponded directly to one of the Gujarati listeners' divisions along H1-H2.)

One might ask whether pathologically-disordered phonation and linguistically-relevant phonation are perceptually the same. Recall that the pathologically-disordered stimuli were specifically selected to resemble the linguistic phonation contrasts in Mazatec so that a direct comparison of the experiments could be made. We saw that Gujarati listeners treated the pathologically-disordered stimuli the same way that they treated the Mazatec stimuli. (With both the Mazatec and pathologically-disordered stimuli, Gujarati listeners based their judgments solely on H1-H2 and mapped the stimuli into three clusters with similar H1-H2 values across the two experiments.) However, both the English and Spanish listeners treated the pathologically-disordered voices and the Mazatec stimuli in slightly different ways. While both English and Spanish listeners grouped the Mazatec and pathologically-disordered stimuli into clusters with similar values for the acoustic measures, the listeners differed in how they weighed the importance of the different dimensions. For the Mazatec stimuli, English and Spanish listeners both weighed dimension 1 over dimension 2, while for the pathologically-disordered stimuli, within each of the languages, listeners differed on the relative importance of the dimensions.

We will compare these results to the results from the two previous chapters in the conclusion.

Chapter 6: Conclusion

6.1 Summary of Results

This dissertation investigated the following questions:

> 1) How does linguistic experience affect listeners' perception of phonation?
> 2) What acoustic properties correlate with listeners' perception?
> 3) Are listeners who are sensitive to linguistically-relevant phonation differences more sensitive to pathologically-disordered phonation?

Based on previous literature on phonation and speech perception, it was hypothesized that (1) Gujarati listeners, who have a contrast between breathy and modal phonation in their native language, would be successful in distinguishing modal and breathy vowels with uniform judgments across listeners at least when the phonation of the stimuli were Gujarati-like, and (2) Spanish listeners would perceive phonation differences poorly because they do not have a category distinction in their native language on which to base their perceptions. However, for English listeners, the appropriate hypothesis was less clear. Studies on the perception of allophonic variation, such as Beddor and Strange (1982) on the perception of vowel nasalization, suggest that listeners can distinguish non-contrastive, allophonic stimulus differences; if this ability is a general one, then English listeners should be sensitive to breathy phonation because it occurs allophonically in English. However, results of studies on the perception of pathologically-disordered phonation suggested that English listeners would distinguish phonations poorly (e.g. Kreiman et al. 1990; Kreiman et al. 1992; Kreiman et al. 1993; Kreiman and Gerratt 1996, etc.).

With regards to question two, we know that listeners place importance on different acoustic cues depending on their language background. Based on Fisher-Jørgensen (1967) and Bickley (1982), it was predicted that H1-H2 would be the most salient acoustic property to Gujarati listeners because this correlates with their production of breathy vowels. For English listeners, it was predicted that they would also use H1-H2 (because this correlates with their production of allophonic breathiness), as well as CPP (based on previous research on the perception of phonation (e.g., Klatt and Klatt 1990)). For Spanish listeners, their lack of a vocalic phonation contrast makes it difficult to make a prediction about which cue(s) they might try to use. But, in the absence of any guidance from their language, it was expected that Spanish listeners would rely on the largest auditory difference within the stimuli set.

For question three, it was hypothesized that, based on the breathy phonation category in their own language, Gujaratis would be more consistent in their judgments of pathologically-disordered phonations, while Spanish listeners would lack inter-subject consistency due to their lack of a breathy category. Studies on the perception of pathologically-disordered phonation by English listeners suggested that they would also distinguish phonations poorly.

To answer these questions, English, Spanish, and Gujarati listeners were asked to participate in five tasks: a free-sort task using linguistically-relevant phonation (Chapter 3), three similarity-rating tasks using linguistically-relevant phonation (Chapter 4), and a fourth similarity-rating task using pathologically-disordered phonation (Chapter 5). In the next section, I will summarize the results of these tasks.

6.1.1 Free-sort task with multi-language stimuli

In the results of the free-sort task presented in Chapter 3, we saw that Gujarati listeners were more consistent and did better at sorting the stimuli than either English or Spanish listeners. As predicted, Gujarati listeners' judgments were strongly correlated with H1-H2. Results for Spanish and English listeners showed that they were not very consistent in their judgments, nor could they correctly sort modal from breathy phonation. Despite the allophonic breathiness in English, listeners were neither more consistent nor more accurate than speakers of a language with only modal phonation (Spanish). However, for both English and Spanish listeners, there were patterns in their judgments. Spanish listeners' sorting of the stimuli was significantly correlated with H1-H2 and H1-A1, while English listeners' sorting of the stimuli was significantly correlated with H1-H2.

Although CPP was the most successful acoustic measure (of the eight measures tested) in distinguishing breathy from modal phonation in the stimulus set, there was not a significant relationship between this measure and judgments of any of the three listener groups. The question arises as to why listeners did not use CPP. Perhaps the physical differences among the particular stimuli used in this experiment were not great enough along this dimension. Any arbitrarily small difference in the CPP values of the stimuli could be used by the discriminant analysis to sort the tokens successfully, but it is possible that human listeners have some auditory threshold that potentially limits the usefulness of this parameter. The CPP of the stimuli (Figure 4) range from .5- 4 dB. Perhaps listeners are not sensitive to these small noise differences; however, the just noticeable difference for CPP is currently unknown. We will return to this point after discussing the similarity-rating tasks.

6.1.2 Similarity-rating task with Mazatec stimuli

The results of the three separate similarity-rating tasks using Mazatec stimuli were presented in Chapter 4. Results showed that, across the three similarity-rating tasks, results were similar within each of the three listener groups.

Across the three similarity-rating tasks, English listeners used H1-H2 and CPP to make their judgments, and consistently weighted H1-H2 greater than CPP. In terms of H1-H2, English listeners always mapped the stimuli into roughly two clusters, one with an H1-H2 value of less than 9 dB and the other with an H1-H2 value of more than 9 dB. Furthermore, along H1-H2, the English listeners' division of the stimuli never corresponded to any gaps in the stimulus set. However, in terms of CPP, judgments were not always as consistent, in that the CPP values of the clusters the English listeners created were different in each of the three experiments. These differences in judgment corresponded to measured acoustic differences in the stimulus sets. For example, in Experiment 2 listeners clustered the stimuli into only two groups (9 dB or greater, and less than or equal to 5 dB), instead of three clusters as in Experiments 1 and 3. One reason the stimuli were clustered into only two groups was that there were not many stimuli with intermediate CPP values (between 6 and 9 dB).

One additional point of interest is that the English listeners did not use CPP when making their judgments in the free-sort task. As previously mentioned, one possible reason for why English listeners used CPP in the similarity-rating tasks, but not in the free-sort task, is because the CPP values of the stimuli used in that experiment were very narrow (from .5- 4 dB). We saw in the free-sort task that 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 was greater (i.e., in the similarity-rating tasks), listeners used the information in a way that allows one to see what counted as a "small difference": when presented with a broader range of CPP values, English listeners divided the stimuli into at least two clusters (one with CPP values less than 5 dB, and the other with CPP values greater than 5 dB).

In all three similarity-rating tasks, Spanish listeners based their judgments on H1-H2 and H1-A1, consistently weighting H1-A1 greater than H1-H2. Across the three experiments, Spanish listeners used H1-A1 in similar ways, always grouping the stimuli into two groups, one with H1-A1 values less than 0 dB, and one with H1-A1 values greater than 0 dB. Also in all three experiments, when relying on H1-A1, the Spanish listeners' division of the stimuli roughly corresponded to one of the natural divisions in the stimulus set. When basing their judgments on H1-H2, Spanish listeners mapped the stimuli in a way just like the English listeners (that is, the stimuli were clustered into roughly two groups, one with an H1-H2 of less than 9 dB and the other with an H1-H2 value of more than 9 dB). When using H1-H2, the Spanish listeners' judgments did not correspond to any natural division in the stimulus set.

Finally, the Gujarati listeners relied on H1-H2, and were consistent in that all 12 listeners weighed H1-H2 as the only significant dimension. In all three experiments, the Gujarati listeners' perceptual mapping of the stimuli was clustered into three groups, with very similar H1-H2 values across experiments. These clusters roughly corresponded to stimuli with H1-H2 values of 1-5 dB, 5-13 dB, and greater than 15 dB. Even though Gujarati only has two phonation categories, the Gujaratis consistently mapped the stimuli into three clusters. There appears to be a ceiling for acceptable breathy phonation in Gujarati: anything with H1-H2 values in the range of 1-5 dB is considered "modal", stimuli with H1-H2 values of approximately 5-13 dB are "breathy", and stimuli with H1-H2 values greater than 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'. However, in the present study, when forced to sort the stimuli into two groups, as in the free-sort task, listeners placed stimuli with H1-H2 values greater than 15 dB in the same box as stimuli with H1-H2 values of 5-13 dB.

Furthermore, the perceptual mapping of the stimuli by the Gujaratis, though categorized, did not correspond to the breathy versus modal distinction in Mazatec. First, the Gujaratis always mapped the stimuli into three clusters, despite the fact that there are only two categories in Mazatec. Second, Gujaratis sometimes judged stimuli with two different phonation categories to be perceptually similar. For example, Talker 2 (see Figure 50) produced categorically modal vowels with high H1-H2 values (around 8 dB), but the Gujaratis judged these stimuli to be perceptually more similar to the Mazatec breathy vowels than to modal ones.

The perceptual mapping of the stimuli reflected the breathy versus modal distinction in Gujarati to some extent. In the 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 7-12 dB. There were no stimuli with H1-H2 values higher than 12 dB in the sample measured for this study.

Taken together, the experiments in this study act like a categorical perception task. The free-sort task 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. The similarity-rating task, while not a straightforward categorical perception task, can be interpreted as one. The responses were continuous in nature, but if the responses fall into clusters in the MDS map of the perceptual space, these clusters could be interpreted as categories. Between clusters of stimuli, Gujarati listeners had sharp boundaries, which roughly corresponded to the boundary between breathy and modal phonation in Gujarati. Listeners did not hear stimuli with intermediate H1-H2 values as intermediate in nature, but rather always formed clusters with the stimuli in the similarity-rating task. Furthermore, the Gujaratis also favored the acoustic cue associated with phonation in their native language (H1-H2), even though other acoustic measures (such as CPP) were superior in distinguishing the phonation types of the stimuli used in the experiments.

6.1.3 Similarity-rating task with pathologically-disordered stimuli

The results of the similarity-rating task using pathologically-disordered voices were presented in Chapter 5. These results were very similar to the results of the three similarity-rating tasks presented in Chapter 4. Once again, English listeners based their judgments on two acoustic dimensions, H1-H2 and CPP. When using H1-H2, listeners mapped stimuli into two clusters, one with H1-H2 values of 1-9 dB, and the other with H1-H2 values greater than 9 dB. In addition, English listeners used CPP and, again, mapped the stimuli into roughly two clusters, one with CPP values of -3 to 5 dB, and one with CPP values of 8-15 dB. The division the English listeners made between the clusters along both H1-H2 and CPP roughly corresponded to a division in the stimulus set. However, individual English listeners were not consistent, in that they differed in which dimension they weighed more heavily.

Spanish listeners based their judgments on two acoustic dimensions, H1-A1 and H1-H2. When Spanish listeners based their judgments on H1-A1, the stimuli mapped into roughly two clusters, one with H1-A1 values of -7 to 1 dB, and the other with H1-A1 values of 4-22 dB. In addition, Spanish listeners used H1-H2 to map the stimuli into roughly two clusters, one with H1-H2 values less than 9 dB, and one with values greater than 9 dB. Along the H1-H2 dimension, but not along H1-A1, the division the Spanish listeners made between the clusters roughly corresponded to a division in the stimulus set. However, individual Spanish listeners differed in which dimension they weighed more heavily.

The Gujarati listeners used H1-H2 to make their judgments, and as with the previous similarity-rating tasks, mapped the stimuli into three clusters corresponding to H1-H2 values of 1-3 dB, 5-13 dB, and 15-31 dB. Recall that that there was a gap between the stimulus with a measured H1-H2 value of 12.5 dB and the stimulus with a value of 17.1 dB, a gap which corresponds directly to one of the Gujarati listeners' divisions along H1-H2. In addition, Gujaratis were more consistent than English and Spanish listeners, in that each individual Gujarati listener used H1-H2 to make his/her judgment.

6.2 General discussion

Recall that one of the main purposes of this study was to answer the question of how linguistic experience affects the perception of linguistically-relevant phonation. We saw that speakers of a language that has a phonemic phonation contrast did better at distinguishing the same type of contrast in other languages. However, speakers of a language with an allophonic distinction did no better than speakers of a language with no phonation contrast at all. These results suggest that listeners are not aided by just any distinction, but rather that a particular type of distinction (i.e., a *phonemic* phonation contrast) is crucial for the perception of phonation.

In terms of consistency, we saw that listeners of Gujarati, a language with a phonemic phonation contrast, were more consistent than listeners of a language with either an allophonic contrast (English), or no contrast at all (Spanish). In fact, these latter types of speakers behaved very similarly in terms of consistency, again showing that allophonic variations do not aid much in the perception of phonation.

An additional purpose of this study was to determine which acoustic dimensions listeners would use to make their judgments. Listeners of a language with a phonemic phonation contrast relied solely on the acoustic measure (H1-H2) that is associated with the production of phonation in their own language. Listeners of a language with an allophonic phonation contrast (English) relied weakly on the acoustic measure (H1-H2) that is associated with the production of phonation in their own language. However, in the three similarity-rating tasks, but not in the free-sort task, these listeners also used CPP as the second most important dimension after H1-H2. 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 than the amplitude of the fundamental frequency (H1).

Finally, we saw that speakers of a language without a phonation contrast (Spanish) did rely consistently on two particular acoustic measures (H1-A1 and H1-H2) rather than being completely random in their judgment. However, H1-A1 and H1-H2 were not the best measures in either experiment based on the discriminant analysis. We can ask ourselves, why did the Spanish listeners rely on these two measures? Cross-linguistically, open quotient (H1-H2) is the most common way to produce phonation differences. H1-H2 is a successful measure of phonation in a variety of languages (e.g., Hmong (Huffman 1985), !Xóõ (Ladefoged et al. 1985), Gujarati (Bickley 1981), Chong (Blankenship 1997), Jalapa Mazatec (Blankenship 1997), allophonic variation in American English (Epstein 1999), etc.). Since all healthy speakers are equally able to make all possible articulations, the reason that H1-H2 is favored in so many languages could be a perceptual one. Perhaps, open quotient differences are naturally more salient to human listeners. This would explain why Spanish listeners relied on H1-H2. However, this does not explain why the Spanish listeners used H1-A1, which is not that well represented cross-linguistically. It is worth investigating Spanish connected speech to see how H1-A1 might vary (e.g., does intonation or lexical stress play a role?).

In addition, Spanish and English listeners both used H1-H2 in the same way. Throughout the various experiments in this dissertation, the English and Spanish listeners 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, that the listeners might pick up on. If presented with a smooth range of stimuli, would English and Spanish listeners still use the 9 dB cut-off?

The last question addressed in this study is how listeners treat pathologically-disordered phonations. We saw that listeners of a language with a phonemic phonation contrast behaved more consistently and efficiently (in that they relied on only one measure) than the other listener types used in this study, who disagreed on the relevant importance of the measures that they used. Furthermore, we saw that Gujarati listeners treated the pathologically-disordered stimuli and the Mazatec stimuli in the same way. This provides support in terms of perception for Ladefoged's (1983) statement that one person's voice disorder is another person's phoneme, which is already supported in terms of production (see Epstein 1999).

However, English and Spanish listeners treated the pathologically-disordered voices and the Mazatec stimuli in slightly different ways. The listeners differed in how they weighed the relative importance of the different dimensions. For the Mazatec stimuli, English and Spanish listeners both weighed dimension 1 over dimension 2, while for the pathologically-disordered stimuli, within each of the languages, listeners differed on the relative importance of the dimensions. This suggests that, to English and Spanish listeners, pathologically-disordered phonations are subtly different from linguistically-relevant phonations. The results of the study presented here suggest that, while Ladefoged's (1983) observation that "one person's voice disorder is another person's phoneme" holds true for listeners with a phonemic phonation contrast (Gujarati), the issue is more complex for listeners with allophonic distinctions (English) or no distinction at all (Spanish).

A potential concern for this study is that the Spanish and Gujarati listeners who participated in this experiment were all bilingual (to one degree or another). However, results indicate that these listeners were not using English to make their judgments. If these listeners were basing their judgments solely on English, they would have performed in the same way as the English listeners, yet this is not the case. While the Gujaratis did base their judgments on H1-H2, just as the English listeners did, the Gujaratis used H1-H2 in a different way (e.g., they created three groups from the stimuli rather than two in the similarity-rating task, they had greater per-pair consistency in the free-sort task, etc.). Moreover, while Spanish and English listeners did indeed perform very similarly, they differed in a few crucial ways. Notably, Spanish listeners relied on H1-A1 and not CPP, and while they relied on H1-H2 in the three similarity-rating tasks with Mazatec stimuli, it was not their most important dimension, as it was for the English listeners. Ultimately, however, we still do not know how monolingual Spanish and/or Gujarati speakers would perform in tasks on the perception of phonation. The next step would be to repeat this study using only monolingual speakers.

6.3 Future research

In this section, I will discuss three areas of possible future research:

1) The possible phonation contrasts in Mexican Spanish,

2) The replication of these results with synthesized stimuli,

3) The replication of these results with more natural stimuli.

In term of the Spanish listeners, it was surprising that they performed so similarly to the English listeners. However, we know that the Spanish listeners, who all spoke English as a second language, were not simply listening as English listeners because they used H1-A1 and not CPP. In addition, they were slightly better at distinguishing modal from breathy vowels than the English listeners in the free-sort task. These findings raise the question as to whether Spanish does have some non-modal phonation. For the purpose of experiments like this one, it would be useful to conduct a study on the production of phonation in Spanish on the same scale as the one done for English.

Regarding the second point, we know that synthesized speech has proven to be useful in perception tasks, especially because it allows for more control over the stimuli. However, prior to this experiment, it was not clear what parameters should be synthesized, and there is no synthesizer that would allow this degree of control over phonation parameters. Thus, the second avenue for future research is to try to replicate this study using synthesized stimuli. More specifically, synthesized stimuli would allow us to test one acoustic property at a time, and to create a more even continuum of stimuli along a given acoustic dimensions. This future study would allow us to determine a more precise perceptual cut-off between breathy and modal phonation along a given dimension, for a given language.

On the other hand, another avenue for future research would be to use stimuli with more natural variation (that is, stimuli with natural f0 and duration). The question arises, would Gujaratis still listen to only H1-H2 if the stimuli had nothing normalized? That is, would natural variation (i.e. f0, duration, etc.) in the stimuli distract the Gujarati listeners or would they be able to ignore these other factors and only listen to H1-H2?

In conclusion, we have seen that while linguistic experience can play a role in the perception of phonation, different types of experience are not equivalent. Even though a phonemic contrast in phonation aids listeners, an allophonic distinction does little to benefit them.

References

Abramson, A., T. Thongkum, and P.W. Nye. 2004. Voice register in Suai (Kuai): An Analysis of Perceptual and Acoustic Data. *Phonetica* 61, 147-71.

Andrade, A. 2003. *Intonational phonology of alteño Spanish*. UCLA, M.A. thesis.

Andruski, J., and M. Ratliff. 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, MIT, 73-93.

Beddor, P.S., and W. Strange. 1982. Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustic Society of America* 71, 1551-1561.

Best, C.T., G.W. McRoberts, and N.M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14, 345-60.

Blankenship, B. 1997. *The Time Course of Breathiness and Laryngealization in Vowels*. UCLA, Ph.D. dissertation..

Bohn, O-S. and J.E. Flege. 1990. Perception and production of new vowel categories by adult second language learners. In *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquistion of Second Language Speech (University of Amsterdam, 9-12 April 1990)*. Leather, J., and A. James, editors., University of Amsterdam.

Canfield, D. L. 1981. *Spanish Pronunciation in the Americas.* Chicago: The University of Chicago Press.

Dilley, L., S. Shattuck-Hufnagel, and M. Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24, 423-444.

Elman, J.L., R.L. Diehl, and S.E. Buchwald. 1977. Perceptual switching in bilinguals. *Journal of the Acoustic Society of America* 62, 971-74.

Epstein, M. 1999. *A Comparison of Linguistic and Pathological Breathiness Using the LF Model*. UCLA, M.A. thesis.

Epstein, M. 2002. *Voice Quality and Prosody in English*. UCLA, Ph.D. Dissertation.

Esposito, C.M. 2003. *Phonation in Santa Ana del Valle Zapotec*. UCLA, M.A. thesis.

Esposito, C.M. 2004. Santa Ana del Valle Zapotec Phonation. *UCLA Working Papers in Phonetics* 103.

Esposito, C.M. 2005. An Acoustic and Electroglottographic Study of Phonation in Santa Ana del Valle Zapotec. Poster presented at the 79th meeting of the Linguistic Society of America, San Francisco, CA.

Esposito, C.M. 2006. Pilot of for the effects of linguistic experience on the perception of phonation. UCLA unpublished manuscript.

Fischer-Jørgensen, E. 1967. Phonetic analysis of breathy (murmured) vowels. *Indian Linguistics* 28, 71-139.

Forrest, K., G. Weismer, P. Milenkovic, and R.N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84, 115-123.

Fox, R., J. Flege, and M. Munro. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America* 97, 2540-2551.

Gandour, J., and R. Harshman. 1978. Cross-language differences in tone perception: A multidimensional scaling investigation. *Language and Speech* 21-33.

Gerfen, C., and K. Baker. 2005. The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics* 33, 311-334.

Gordon, M., and P. Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383-406.

Granqvist, S., and L. Eng. 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., M.W. Post, and D.L. Payne. 2004. Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics* 32, 517-542.

Hillenbrand, J.M., R.A. Cleveland, and R.L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37, 769-778.

Holmberg, E.B., R.E. Hillman, J.S. Perkell, P.C. Guiod, and S.L. Goldman. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research* 38, 1212-1223.

Huffman, M.K. 1985. Measures of phonation type in Hmong. *UCLA Working Papers in Phonetics* 61.

Huffman, M.K. 1987. Measures of phonation type in Hmong. *Journal of the Acoustical Society of America* 81, 495-504.

Ingram, J.C.L., and S-G. Park. 1998. Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America* 103, 1161-1174.

Iverson, P., and P. K. Kuhl. 1996. Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America* 99, 1130–1140.

Johnson, K. 2003. *Acoustic and auditory phonetics*. Oxford: Blackwell Publishing.

Kempster, G. B., D.J. Kistler, and J. Hillenbrand. 1991. Multidimensional scaling analysis of dysphonia in two speaker groups. *Journal of Speech and Hearing Research* 34, 534–543.

Kirk, P.L., J. Ladefoged, and P. Ladefoged. 1993. Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. In *American Indian Linguistics and ethnography in honor of Lawrence C. Thompson.* A. Mattina and T. Montler, editors., Missoula, MT: Universtiy of Montana Press.

Klatt, D.H., and L.C. Klatt. 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., B.R. Gerratt, and K. Precoda. 1990. Listener experience and perception of voice quality. *Journal of Speech and Hearing Research* 33, 103-115.

Kreiman, J., B.R. Gerratt, N. Antonanzas-Barroso, and G.S. Berke. 1993. Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research* 36, 14-20.

Kreiman, J., B.R. Gerratt, and G.S. Berke.1994. The multidimensional nature of pathological voice quality. *Journal of the Acoustical Society of America* 96(3), 1291-1301.

Kreiman, J. and B.R. Gerratt. 1996. The perceptual structure of pathologic voice quality. *Journal of the Acoustical Society of America* 100(3), 1787-1795.

Kreiman, J., and B. R. Gerratt. 1999. Measuring vocal quality. In *Handbook of voice quality measurement*. Kent R., and M. J. Ball, editors. San Diego: Singular Publishing Group.

Kreiman, J., and B. R. Gerratt. 2000. Sources of listener disagreement in voice quality assessment. *Journal of the Acoustical Society of America* 108, 1867-1879.

Kreiman, J., and B. R. Gerratt. 2006. Predicting perceptual salience. unpublished manuscript.

Ladefoged, P. 1971 *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

Ladefoged, P. 1983. The linguistic use of different phonation types. In *Vocal Fold Physiology: Comtemporary Research and Clinical Issues.* Bless D., and J. Abbs, editors., San Diego: College-Hill Press.

Ladefoged, P., and N. Antoñanzas-Barroso. 1985. Computer measured of breathy voice quality. *UCLA Working Papers* 61.

Ladefoged, P., and I. Maddieson. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell Publishers.

Laver, J. 1981. *Phonetic Description of Voice Quality.* Cambridge: University of Cambridge Press.

Lee, T. 1983. An acoustical study of the register distinction in Mon. *UCLA Working Papers in Phonetics* 57, 79-96

Liberman, A.M., K.S. Harris, H.S. Hoffman, and B.C. Griffith. 1957. The discrimination of speech sounds with and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358-368.

Lofqvist, A., and R.S. McGowan. 1992. Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics* 20, 93-110.

Logan, J.S. and J.S. Pruitt. 1995. Methodological issues in training listeners to perceive nonnative phonemes. In *Speech perception and linguistic experience: Issues in cross-language research.* Strange, W., editor. Maryland: York Press Inc.

Mazaudon, M. 2003. Tamang. In *Sino-Tibetan Languages.* LaPolla R. and G. Thurgood, editors. London / New York: Routledge.

Munro, P. and F.H. Lopez, with O. Mendez, R. Garcia, and M.Galant. 1999. *Di'syonaary X:tee'n Diizh Sah Sann Lu'uc (San Lucas Quiaviní Dictionary/ Diccionario Zapoteco de San Lucas Quiaviní)*. Los Angeles: Chicano Studies Research Center Publications, UCLA.

Nair, U. 1979 *Gujarati phonetic reader*. Mysore: Central Institute of Indian Languages.

Ní Chasaide. A., and C. Gobl. 1997. Voice Source Variation. In *The Handbook of Phonetic Sciences.* Hardcastle W.J., and J Laver, editors, Oxford Blackwell Publishers.

Pierrehumbert, J. and D. Talkin. 1992. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Docherty G.J. and D.R. Ladd, editors. Cambridge: Cambridge University Press.

Polka, L. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics* 52, 37-52.

Rakerd, B., and R. R. Verbrugge. 1985. Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels. *Journal of the Acoustical Society of America* 77, 296–301.

Silverman, D., B. Blankenship, P. Kirk, and P. Ladefoged, 1995. Phonetic structures of Jalapa Mazatec. *Anthropological Linguistics* 37, 70-88.

Silverman, D. 1997. Laryngeal complexity on Otomanguean vowels. *Phonology* 14, 235-62.

Shrivastav, R., and C.M. Sapienza. 2006. Some difference limens for the perception of breathiness. *Journal of the Acoustical Society of America, 120, 416-423.*

Stevens, K.N. 1977. Physics of laryngeal behavior and larynx modes. *Phonetica* 34, 264-279.

Stevens, Kenneth and H. Hanson. 1995. Classification of glottal vibration from acoustic measurements. In *Voice Quality Control*. Fujimura O., and M. Hirano, editors. San Diego: Singular Publishing Group, Inc.

Strange, W, and S. Dittman. 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics* 36*,* 131-145.

Strange, W., T.R. Edman, and J.J. Jenkins, 1979. Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology: Human Perception and Performance* 5, 643-656.

Tees, R.C., and J. Werker. 1984. Perceptual flexibility: maintenance or recovery of ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38, 579-590.

Terbeek, D. 1977. Cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics* 37.

Thongkum, T. 1987. Phonation types in Mon-Khmer languages. *UCLA Working Papers in Phonetics* 67.

Upperman, G. Changing Pitch with PSOLA for Voice Conversion. http://cnx.org/content/m12474/latest/

Wayland, R and A. Jongman. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics* 31, 181-201.

Weiner, F. F., and S. Singh. 1974. Multidimensional analysis of choice reaction time judgments on pairs of English fricatives. *Journal of Experimental Psychology* 102, 615–620.

Werker, J.F., J.H.V. Gilbert, K. Humphrey, and R.C. Tees. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52, 349-55.

Wilson, S. 2005. Script for free-sort experiment. http://www.linguistics.ucla. edu/ faciliti/facilities/perception/matlab.html

Yamada R., and Y. Tohkura. 1992a. Perception of American English /r/ and /l/ by native speakers of Japanese. In *Speech Perception, Production and Linguistic Structure.* Tohkura Y., E. Vatikiotis-Bateson, and Y. Sagisaka, editors. Tokyo: OHM Publishing Co. Ltd.

Yamada R., and Y. Tohkura 1992b. The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. In *Speech Perception, Production and Linguistic Structure*. Tohkura, Y., E. Vatikiotis-Bateson, and Y. Sagisaka, editors. Tokyo: OHM Publishing Co. Ltd.