The effects of linguistic experience on the perception of phonation

Christina M. Esposito

UCLA Department of Linguistics

esposito@humnet.ucla.edu

Introduction

- Most of the work done on linguistically-relevant phonation involves experimental descriptions of the production of phonation in particular languages.

- Examples:
  - Gujarati (Fischer-Jørgensen 1967)
  - Mazatec (Silverman et al. 1995)
  - Zapotec (Esposito 2003)

Introduction

- This research has been expanded to include specific properties of phonation, such as:

    - the localization (within a vowel) of non-modal phonation (Silverman 1997)

    - the duration of non-modal phonation (Blankenship 1997)

    - the interaction between tone and phonation (Maddieson and Hess 1987)

    - the interaction between intonation and phonation (Epstein 2002; Esposito 2003)

Previous Research

- However, little is known about the **perception** of phonation.

- Two small-scale studies do exist: Fischer-Jørgensen (1967) and Bickley (1982)

  - Both showed that **vowels with the highest amplitude of the first harmonic (H1) were consistently judged to be breathy by Gujaratis** (whose language contrasts breathy and modal vowels).

Previous Research

- Most of what is known about the perception of phonation comes from studies where English listeners judge pathologically disordered voice qualities.

    - This is 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).

Acoustic Cues to Phonation Contrasts

- In a language with contrastive phonation, there can be a **wide variety of productions:**

  - Example: Zapotec– phonation contrasts are produced by the speed of vocal fold closure for men, and the degree of contact for women (Esposito 2003)

- Thus, **listeners must be able to associate one or more of these acoustic properties with a particular contrast**.

- Studies have shown that **particular cues are more salient to certain listening populations**:

  - Ladefoged and Antoñanzas-Barroso (1985) showed that **spectral tilt** was the more **salient** perceptual cue to !Xóõ breathy vowels **for trained American English listeners**.

    (In !Xóõ, a breathy vowel can be distinguished from a modal one by either a change in spectral tilt or the addition of noise, depending on the speaker.)

Current Study

- This study will address the following questions:

> **Does linguistic experience affect one's perception of phonation?**

Hypothesized result: Speakers of languages with phonation contrasts will perform more consistently across subjects than those without such contrasts.

> **What acoustic property (or properties) correlates with listeners' perception?**

Hypothesized result: The acoustic correlate(s) of phonation in a given language will be the most salient acoustic property to listeners of this language.

To answers these questions, a perception experiment using three languages that differ in their use of breathy voice (English, Spanish, and Gujarati) was conducted.

Methods

- **Listeners**
  - **English**
    - Phonemically, English has **modally phonated** vowels.
    - However, there is **some allophonic non-modal phonation**:
      - creakiness associated with the ends of sentences and paragraphs (Kreiman 1982)
      - creakiness at the beginning of vowel-initial words due to allophonic glottal stop (Dilley et al. 1996)
      - breathiness on vowels after [h] (Epstein 1999)
      - non-modal phonation on pitch accents and at phrase boundaries (Epstein 2002)

– **Gujarati**

- has **phonemically modal and breathy vowels**
- Gujarati breathy vowels are characterized by greater open quotient.
- The most **successful acoustic measure of phonation in Gujarati is H1-H2**
- Gujarati was chosen since it contrasts modal and breathy vowels, but **does not have tone**.

(Gujarati 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.))

(Fischer-Jørgensen 1967)

- **Spanish** (various Mexican dialects)
  - has **modally phonated vowels**
  - **no** (apparent) **allophonic breathiness**
  - Dialects with potential breathiness were avoided:
    - dialects that replace /x/ with [h]
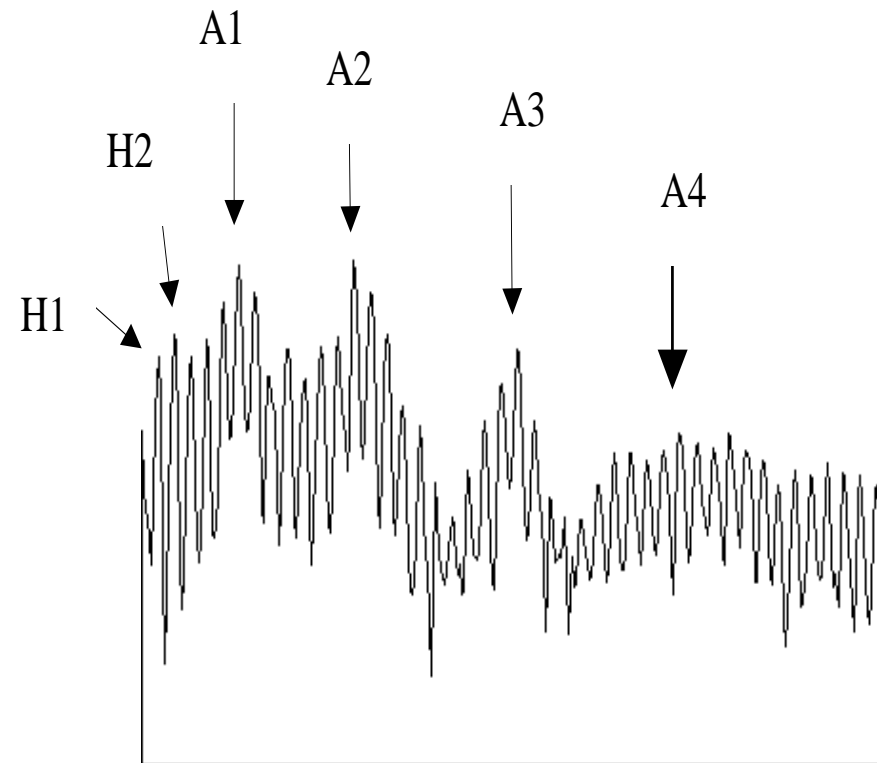   - dialects where /s/ is debuccalized

Stimuli selection

- While synthesized speech has proven to be useful in perception tasks, it is not clear what parameters should be synthesized at this stage in the research on the perception of phonation.

- In place of synthesis, **natural stimuli from languages with breathy and modal vowels were used**.

Measuring phonation

- Measurements
  - The following spectral measures were taken for all the stimuli:
    - **H1-H2** (Bickley 1982)
    - **H1-A3** (Stevens and Hanson 1995; Blankenship 1997)
    - **H1-A1** or **H1-A2** (Ladefoged 1983)
    - the relationship of higher formants to lower ones such as **A2-A3**, **A2-A4**, and **A1-A4** (Klatt and Klatt 1990).
- (Discriminant analysis showed that these measures accounted for 85% of the variance in the stimuli.)

FFT of a modal vowel

Stimuli selection

- **Chong** (Mon-Khmer)
  - Chong has four "tones" which are distinguished by phonation. Tone 3, a falling tone, is produced with breathy phonation.
  - Best acoustic measure of phonation = H1-A2 (Blankenship 1997)
- **Fuzhou** (Sino-Tibetan)
  - There is breathiness associated with the low level tone.
  - Best acoustic measures of phonation = H1-H2 and H1-A2 (Esposito 2005)
- **Green** and **White Hmong** (Hmong-Mien)
  - contrasts 7 tones, each associated with a phonation
  - The low tone is produced with breathy voice.
  - Best acoustic measure of phonation = H1-H2 (Huffman 1985)
- **Mon** (Mon-Khmer)
  - two registers – a high-level tone with modal voice, and a low-level tone with breathy voice
  - Best acoustic measure of phonation = H1-A3 (Lee 1983)

Stimuli selection

- **Santa Ana del Valle Zapotec** (Oto-Manguean)
  - Breathy phonation is always falling tone.
  - Best measures of phonation: H1-H2 for female speakers; H1-A3 for male speakers (Esposito 2003)
- **Tlacolula Zapotec** (Oto-Manguean)
  - Best measure of phonation: H1-H2 (Esposito 2005)
- **Tamang** (Sino-Tibetan)
  - Best measures of phonation: H1-A2 and A2-A3 (Esposito 2005)
- **!Xoo** (Khoisan)
  - Breathy vowels are distinguished from modal ones by open quotient, spectral tilt, and/or noise, depending on the speaker.
  - Best measures of phonation: H1-H2 and H1-A1 (Ladefoged et al. 1985)

Stimuli manipulation

- The stimuli were **composed of breathy and modal vowels excised from real words consisting of an [alveolar stop + vowel].**

- At least 2 breathy and 2 modal vowels were chosen from each language, producing a total of **20 stimuli**.

- **Vowel quality was controlled for, as much as possible.**

- The vowels were cut right after the stop burst, leaving the stop transition intact.

  - This had the effect of :

    - preserving any important phonation cues at the beginning of the vowel

    - avoiding a problematic rise in the amplitude at the onset of the vowel, which could be perceived as a glottal stop

Stimuli manipulation

– Each vowel was normalized to:

- the average value of the duration (250 ms), while preserving the original proportion of breathy to modal phonation

- an f0 of 115-110 Hz using the PSOLA[1] (pitch-synchronous overlap and add) function of PRAAT software

Procedure

- Selecting a procedure
  - While identification tasks are ideal, **the differences between English, Spanish, and Gujarati make it impossible for all three sets of listeners to perform the same identification task**.
  - Possible solutions:

    1. **Train** the English and Spanish speakers on the definitions of "modal" and "breathy"

    Problems:
    - Training would expose listeners to the stimuli prior to the beginning of the experiment.
    - This is difficult since there is no prototypical "breathy" or "modal" phonation to be trained on.

Procedure

- Possible solutions (cont'd):

  2. **Discrimination task**

  Problems:

  - Discrimination tasks do not reflect the influence of the listener's language background as well as identification tasks do.

  - (For example, Ingram and Park (1998), in their experiment on the perception of Australian English /r/ and /l/ by Korean and Japanese speakers, found an effect of the listeners' language background in the identification task, but not the discrimination one.)

Procedure

- The ideal task for this experiment is one that does not make reference to particular category label, while still preserving some elements of an identification task.

  - **Solution = free-sort task**

Free-sort task

- Subjects sort a set of items into groups according to some self-chosen criterion.

- Used successfully in:

  - psychology experiments (Baljko and Hirst 1998)

  - studies on the perception of phonation (Granqvist and Eng 2003).

- Advantages of this method:

  - reference to a category does not need to be made to sort the stimuli

  - task can be performed by all three listening populations

Procedure (free-sort task)

- Implemented in Matlab

- All the stimuli were presented in one trial and arranged in a random order.

- Subjects were told to **sort the stimuli based on "what the voice sounds like" by placing a stimulus (by moving an icon) into one of two boxes, based on perceived similarity.**

- Subjects were **told to play all the stimuli once before sorting any of them.**

- Subjects then picked the stimulus of their choice to sort into one of the boxes.

- This procedure continued until all of the stimuli were sorted.

Schematic representation of the free-sort experiment
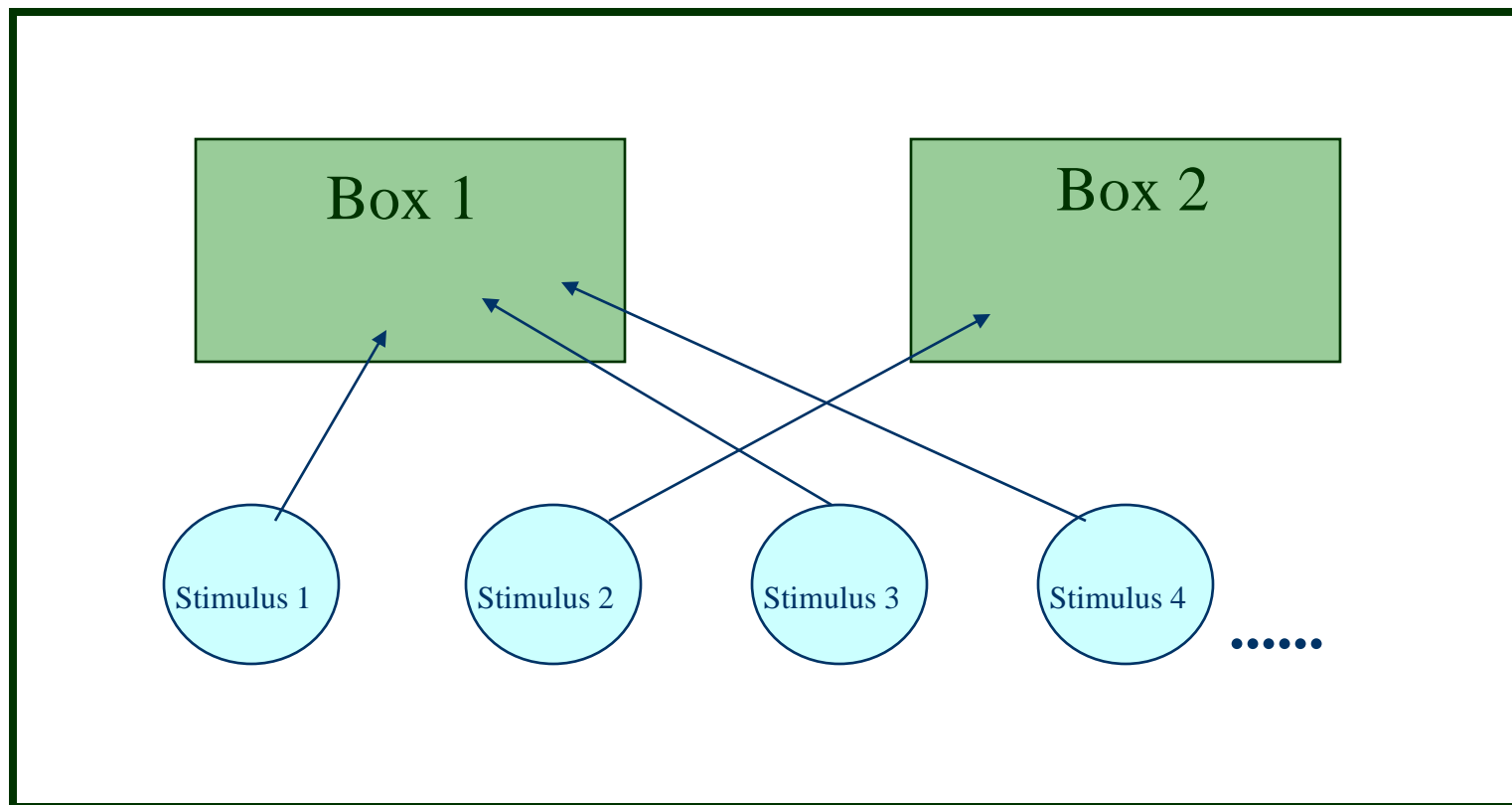
(arrows represent one possible sorting of the stimuli)

Procedure

- Listeners were instructed to **use both boxes** and to **not leave an item alone in a box**.

- Subjects could **listen to the stimuli as often as they liked** and could move a stimulus out of a 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.

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., all possible pairings of stimuli were examined).

- To determine if there was consistency between subjects with regard to particular pairings of the stimuli, the following was calculated:

Per-pair consistency =

(Number of subjects responding "Same" to a pair/Total Number of Subjects) X 100

(2 stimuli in the same box = "same")

- The average per-pair consistency is presented in the chart below.

Results - Consistency

Listeners	Average Consistency
English	46.6% (evenly distributed in the range of 0 to 80%)
Spanish	45% (evenly distributed in the range of 0 to 70%)
Gujarati	⟶ 92.1% (distributed in the range 80 to 100%, with most $\geq$90%)

- **English** - **lack of consistency** within this listener group
- **Spanish** - **lack of consistency** within this listener group
    - English listeners' judgments were not more consistent across listeners, despite the allophonic breathiness found in English.
- **Gujarati** - Gujarati listeners did in fact **behave more uniformly** than English and Spanish listeners.
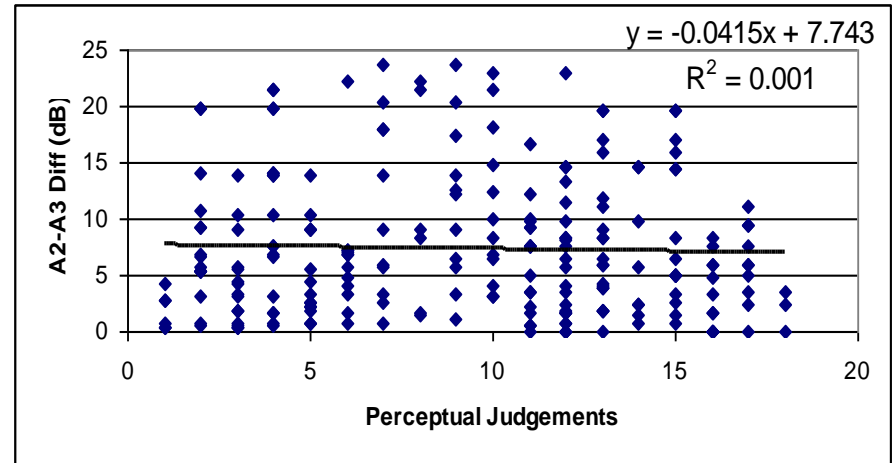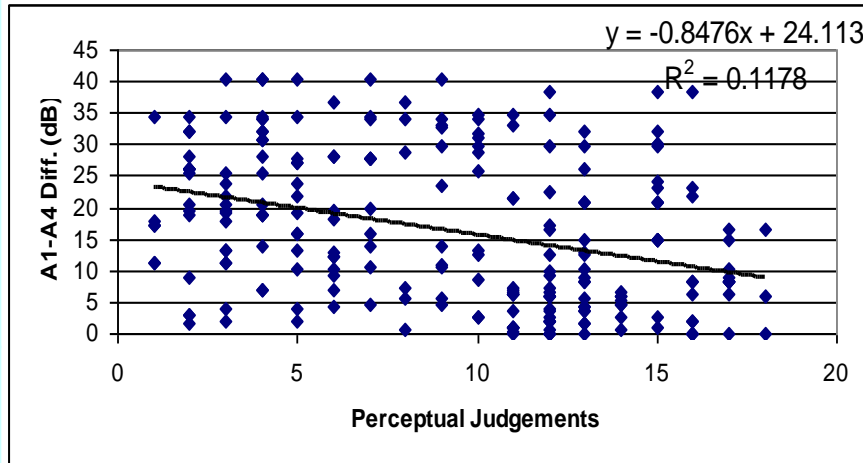
Results – Correlations

- To determine if there was a correlation between the acoustic properties and the perceptual judgments, r^2 was calculated.

- In the following graphs, the x-axis represents the perceptual judgments (that is, the number of listeners who agreed on this pairing of the stimuli). The y-axis represents the difference (in dB) of the two stimuli in question.
  - Example:
    - H1-H2 value for Stimulus A: 20 dB
    - H1-H2 value for Stimulus B: 15 dB
    - H1-H2 difference of Stimuli A and B: 5 dB
  - 7 different measures are presented over 2 slides (per language)
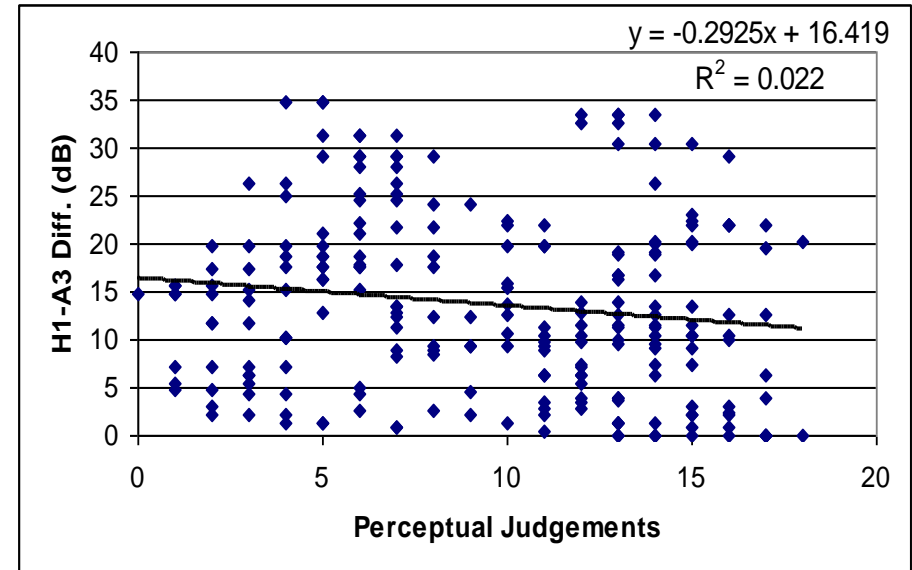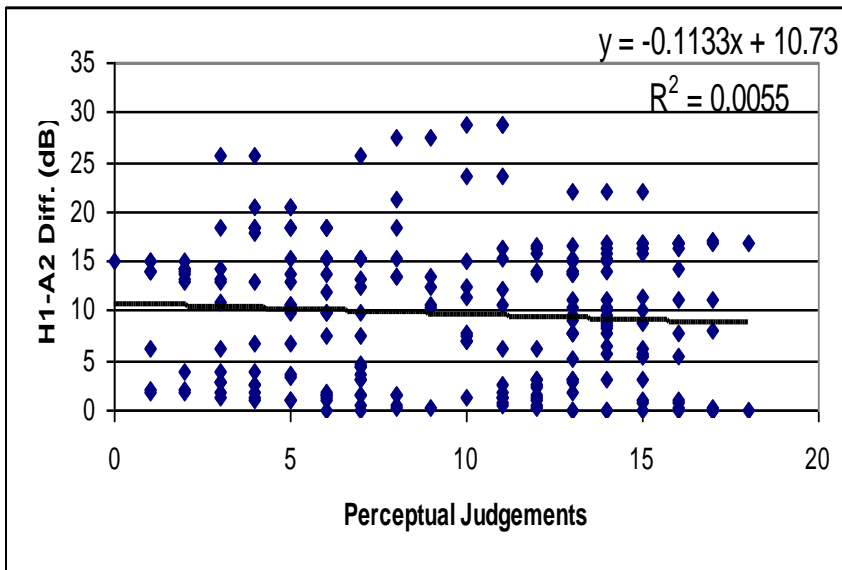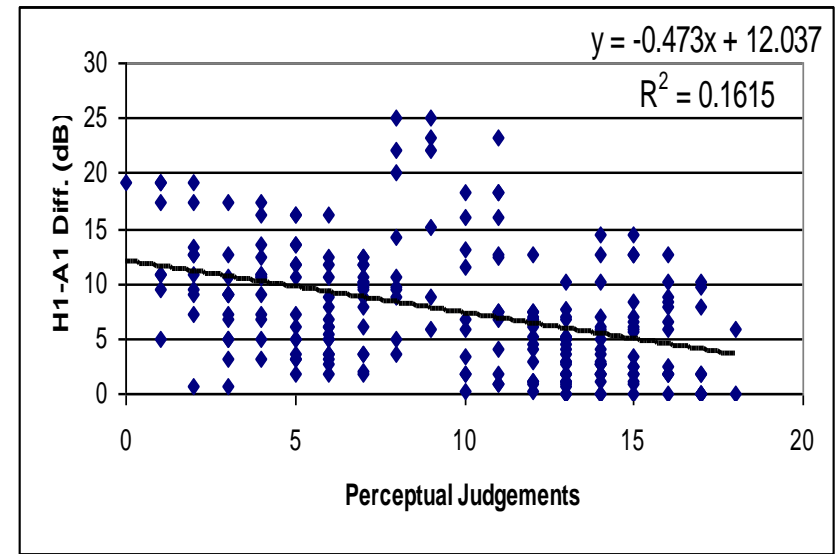- The **smaller the dB difference (for a given measure) between two stimuli, the more likely the stimuli should be paired together**.
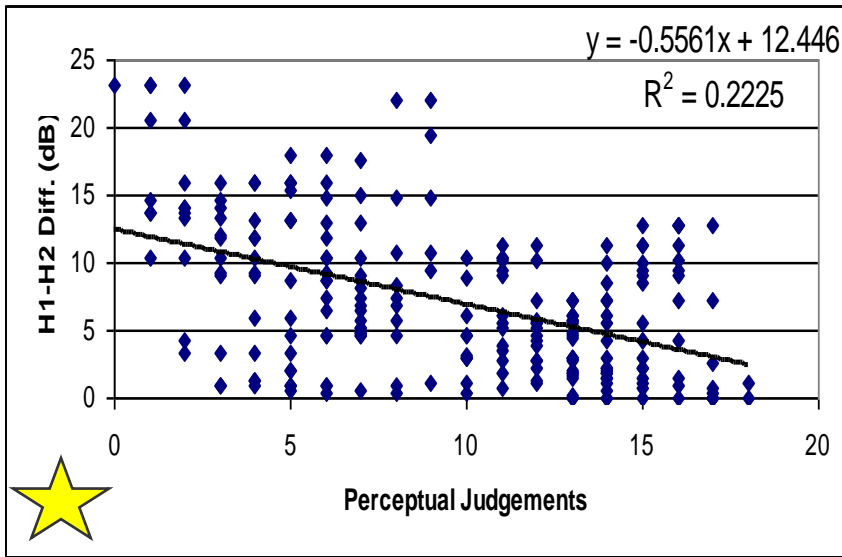
Results – English listeners

$y = -0.4357x + 11.191$

$R^2 = 0.1416$

H1-H2 diff (dB) vs Perceptual Judgements

⭐ =significant

$y = -0.4047x + 11.286$

$R^2 = 0.1226$

H1-A1 Diff. (dB) vs Perceptual Judgements

$y = -0.0706x + 10.306$

$R^2 = 0.0022$

H1-A2 Diff. (dB) vs Perceptual Judgements

$y = -0.248x + 15.933$

$R^2 = 0.0164$

H1-A3 Diff. (dB) vs Perceptual Judgements

Results – English listeners

A1-A4 Diff. (dB) vs Perceptual Judgements

$y = -0.8476x + 24.113$

$R^2 = 0.1178$

A2-A3 Diff (dB) vs Perceptual Judgements

$y = -0.0415x + 7.743$

$R^2 = 0.001$

A2-A4 Diff (dB) vs Perceptual Judgements
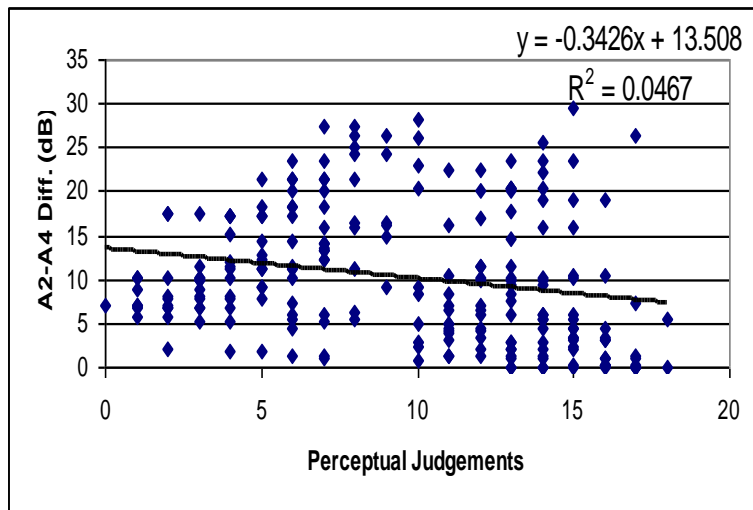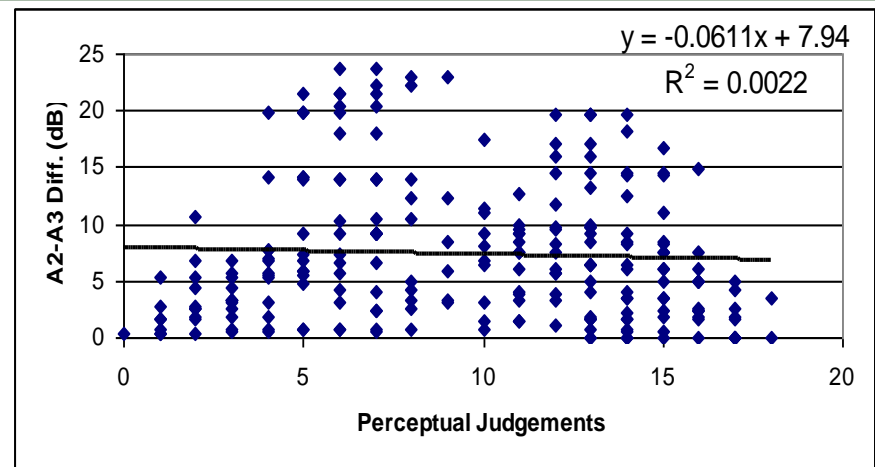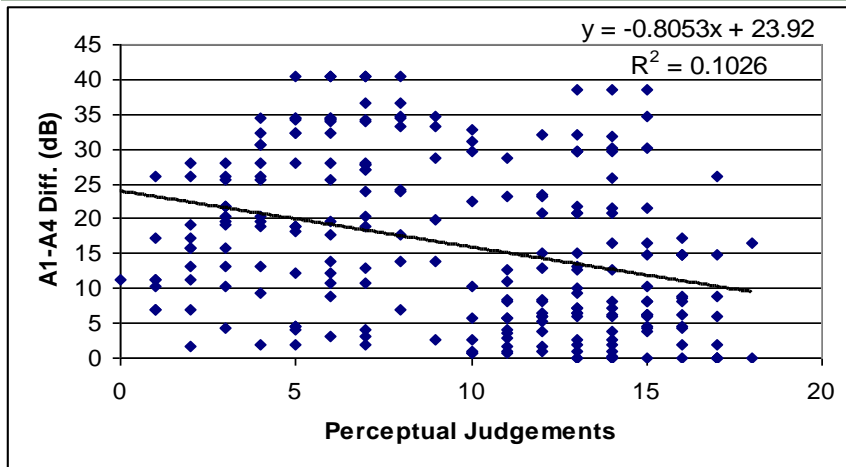
$y = -0.3108x + 13.128$

$R^2 = 0.0399$
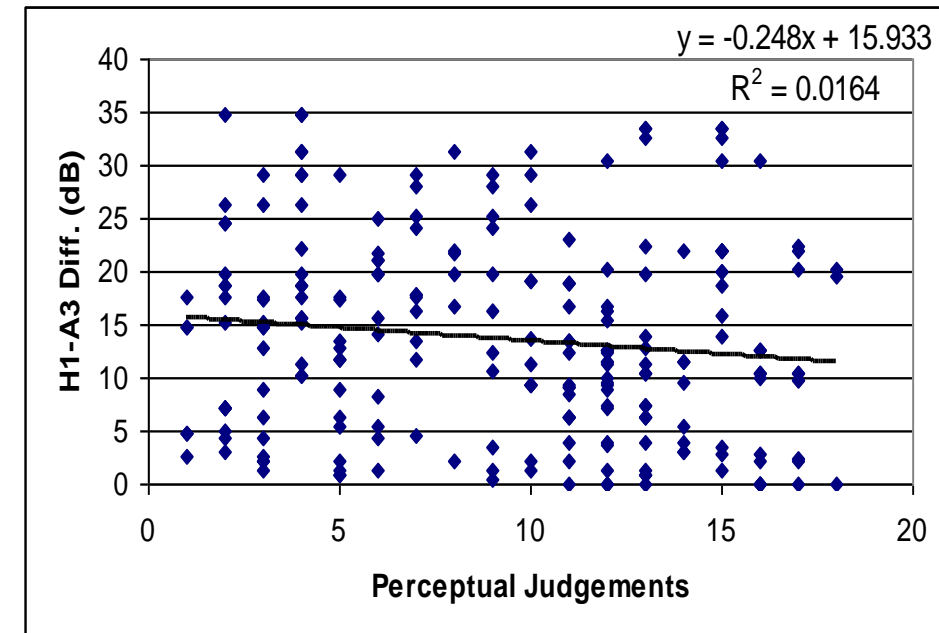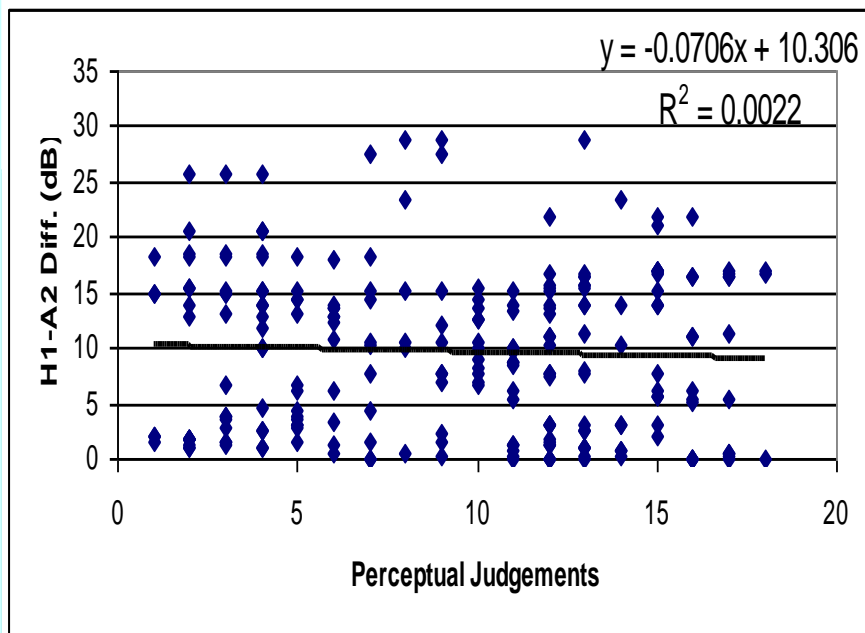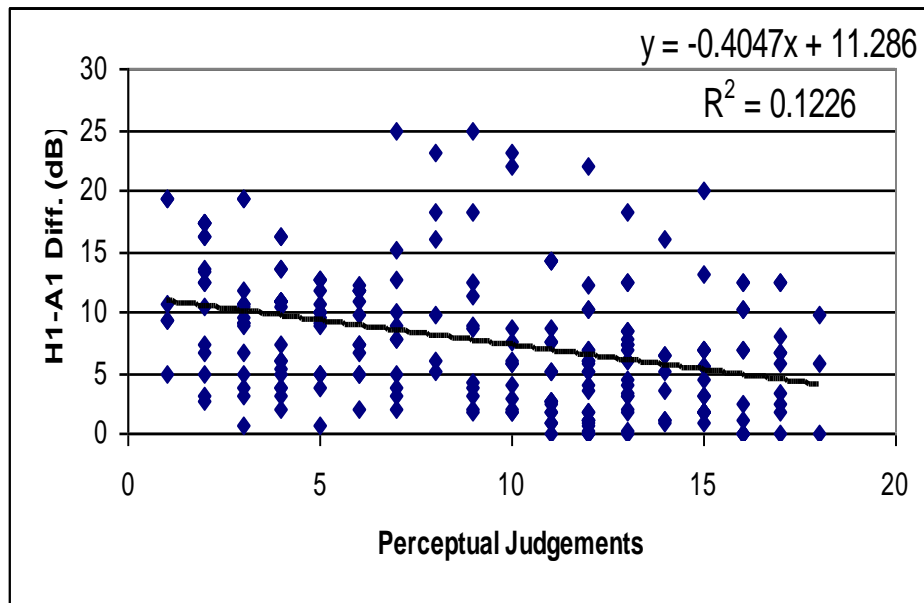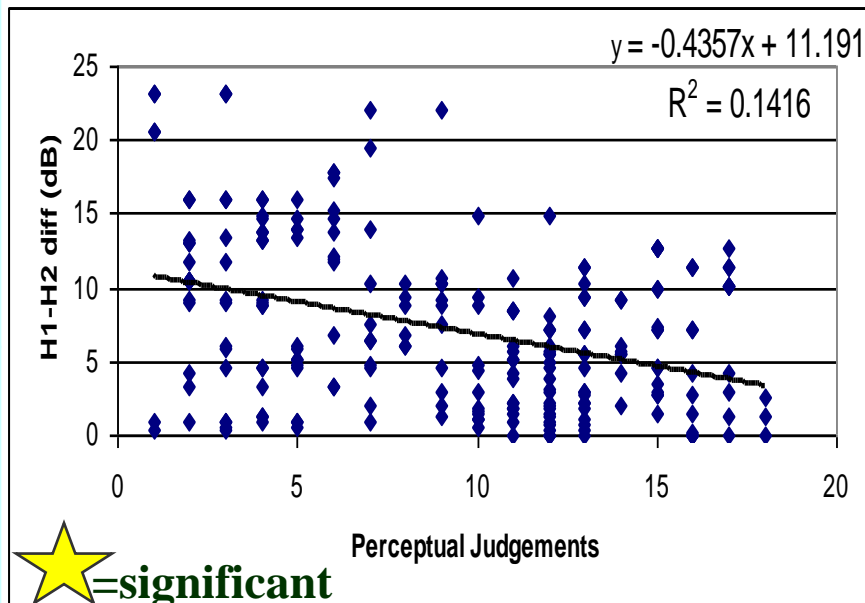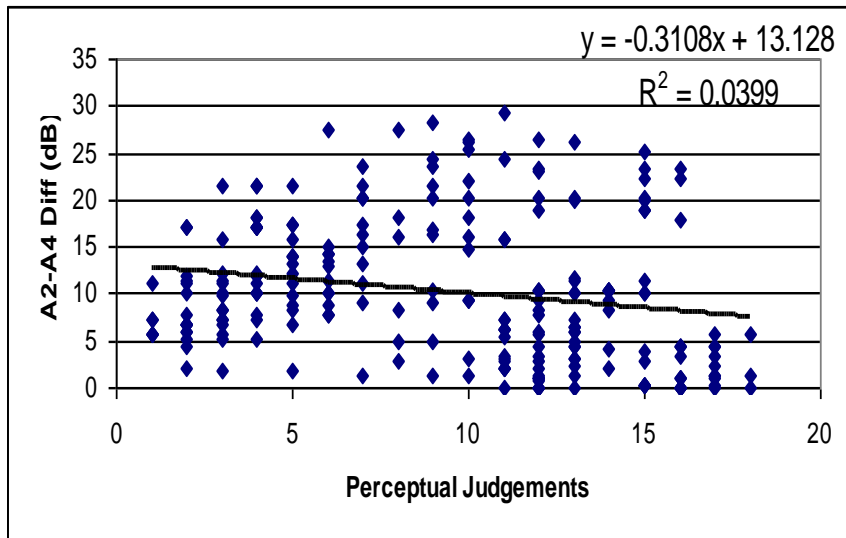
- Overall:
  - **significant but weak relationship between H1-H2** and English listeners' judgments
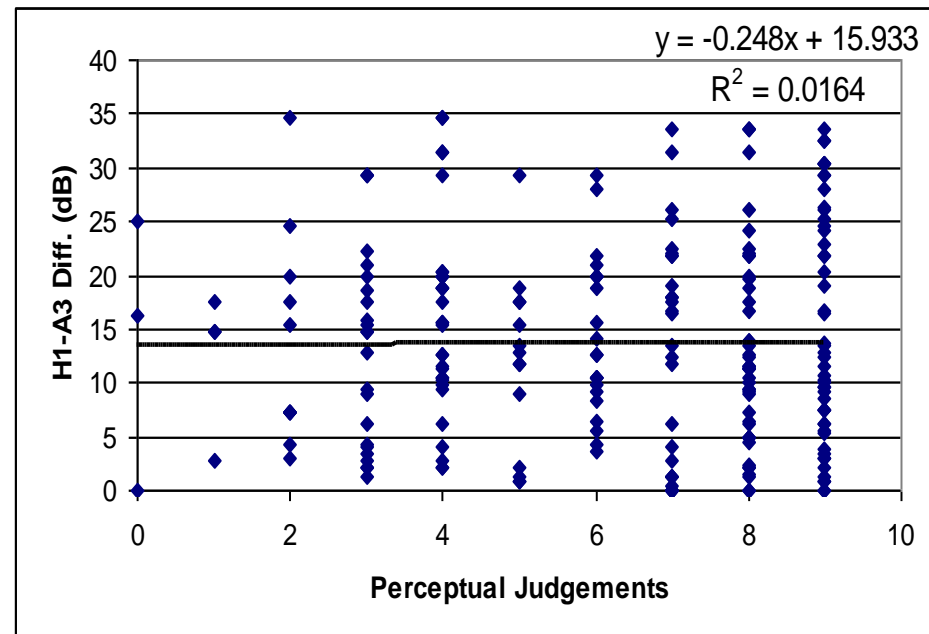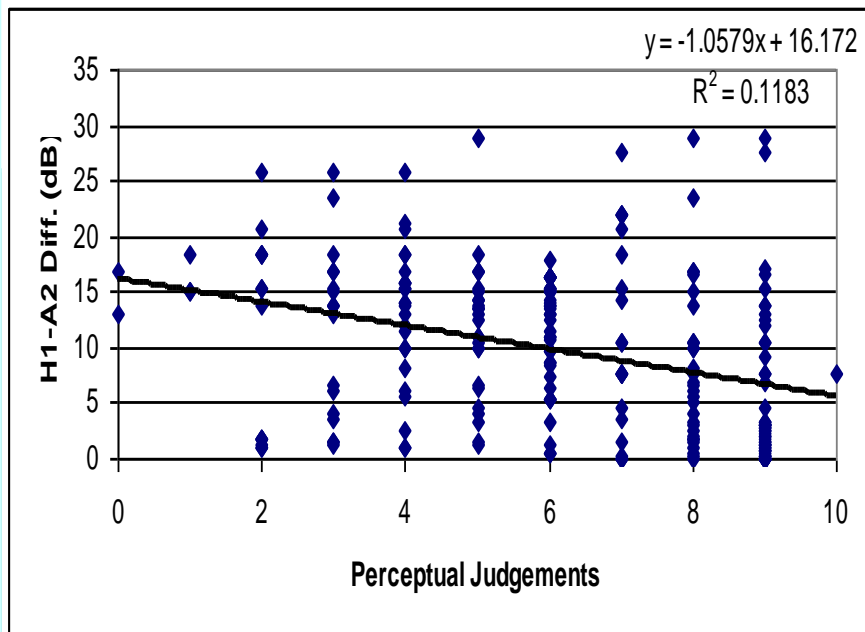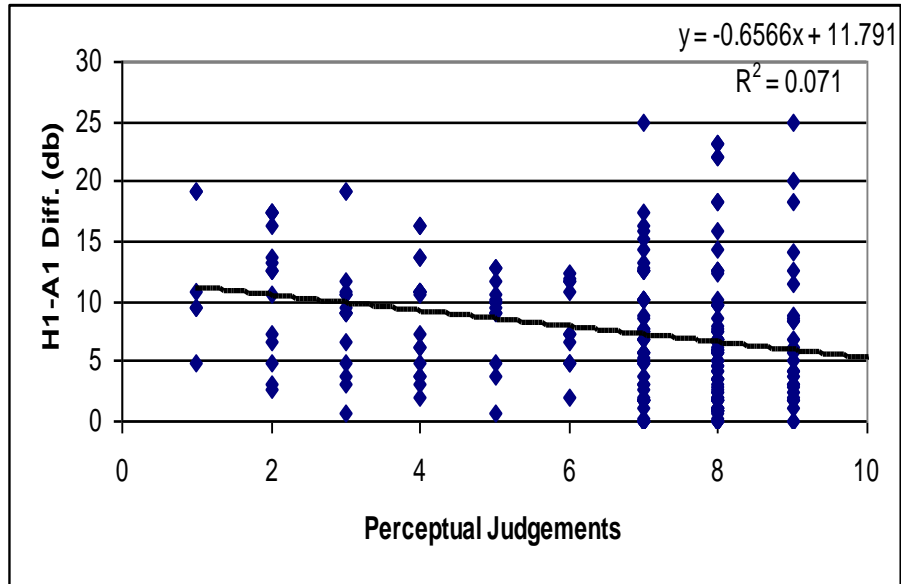 - no significant relationship between the other measures and the English listeners' judgments

Results – Spanish listeners

Results – Spanish listeners

$y = -0.8053x + 23.92$
$R^2 = 0.1026$

A1-A4 Diff. (dB) vs Perceptual Judgements

$y = -0.0611x + 7.94$
$R^2 = 0.0022$

A2-A3 Diff. (dB) vs Perceptual Judgements

$y = -0.3426x + 13.508$
$R^2 = 0.0467$
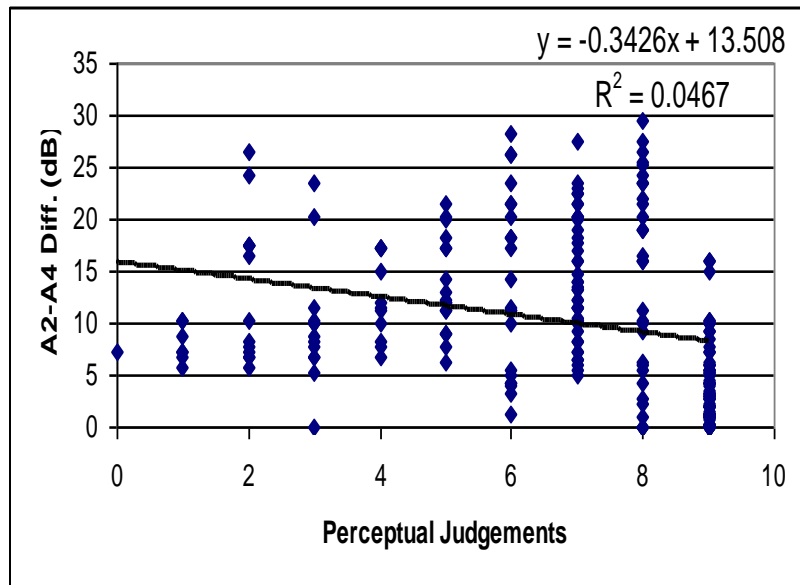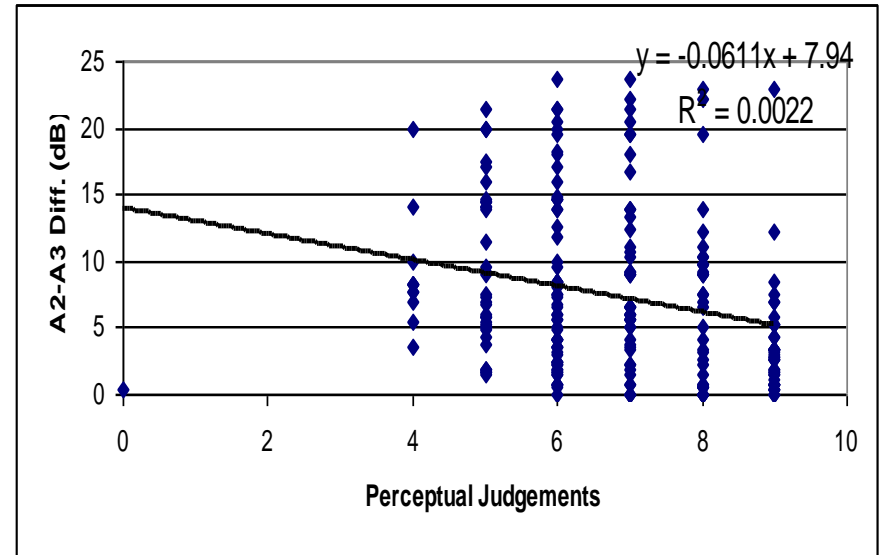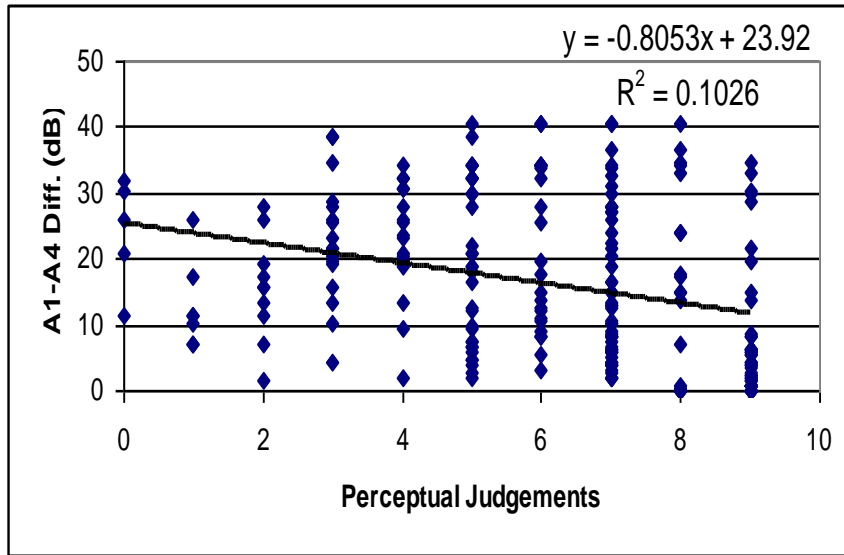
A2-A4 Diff. (dB) vs Perceptual Judgements

- Overall:
  - **significant but weak relationship between H1-H2** and Spanish listeners' judgments
  - no significant relationship between the other measures and the Spanish listeners' judgments

Results – Gujarati listeners

Results – Gujarati listeners

$y = -0.8053x + 23.92$
$R^2 = 0.1026$

$y = -0.0611x + 7.94$
$R^2 = 0.0022$

$y = -0.3426x + 13.508$
$R^2 = 0.0467$
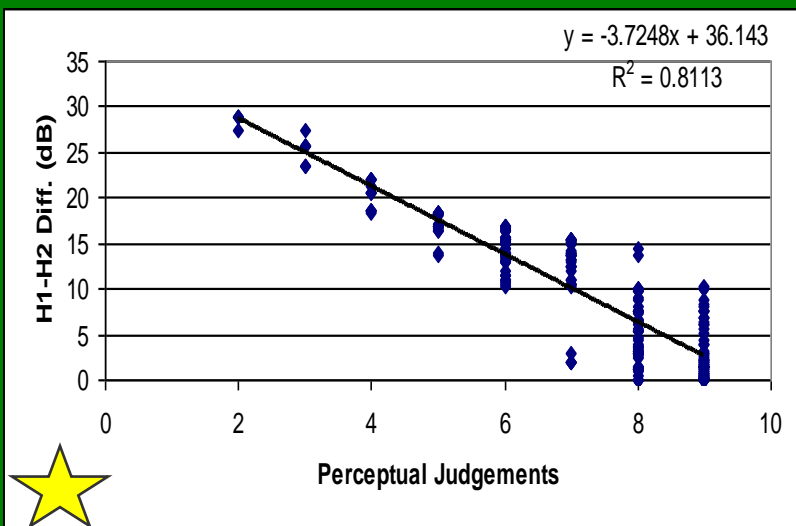
- **Overall:**
  - **strong, significant relationship between H1-H2 and Gujarati listeners' judgments**
  - **The smaller the difference in H1-H2 between two stimuli, the more likely they were to be grouped together.**

Conclusion

- Questions addressed in this study:

  ➢ **Does linguistic experience affect one's perception of phonation?**

  Hypothesized results: Speakers of languages with phonation contrasts will perform more consistently across subjects, than those without such contrasts.

  ➢ **Yes, Gujarati listeners did in fact behave more consistently than English and Spanish listeners.**

  ➢ **Despite the allophonic breathiness in English, listeners' judgments were not more consistent than speakers of a language with only modal phonation (Spanish).**

Conclusion

➢ **What acoustic properties correlate with listeners' perception?**

> Hypothesized results: The acoustic correlate of phonation in given language will be the most salient acoustic property to listeners of this language.

➢ **While there was a significant correlation for all three listener groups, only the Gujarati listeners' judgments were *strongly* correlated with H1-H2, which reflects the production of phonation in their own language.**

Footnotes

[1] Since resynthesizing f0 can create changes in harmonics, the f0 of the stimuli was resynthesized to a value of +/- 40 Hz from the original value. Resynthesizing within this range creates only a > 1 dB change in the harmonics.

Selected References

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

Bickley, C. 1982. Acoustic analysis and perception of breathy vowels. *Speech Communication Group Working Papers*, Research Laboratory of Electronics, MIT 73-93.

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

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. 2003. *Santa Ana del Valle Zapotec Phonation*. UCLA, M.A. Thesis.

Esposito, C. 2005 The effects of linguistics experience on the perception of phonation. unpublished manuscript.

Granqvist and Eng 2003 http://www.speech.kth.se/~svante/Thesis/paperI.pdf

Hanson, H. 1997. Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101: 446-481.

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 Park S-G. 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.

Kreiman, J. 1982. Perception of sentences and paragraph boundaries in natural conversation. *Journal of Phonetics* 10:163-175.

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

Pierrehumbert, J. and Talkin D. 1992. Lenition of /h/ and glottal stop. In G.J. Docherty and D.R. Ladd, eds. *Papers in Laboratory Phonology II: Gesture,Segment, Prosody*. Cambridge University Press: Cambridge, 90-117.