Lifting the Curtain on the Wizard of Oz: Biased Voice-Based Impressions of Speaker Size

Drew Rendall, John R. Vokey, Christie Nemeth

(2007)

Presented by Dasha Gluhareva

Which speaker is larger?

1.

2.

1.

2.

Background

- Listeners make personal evaluations and attributions of speakers based on nonlinguistic voice cues (Giles & Powsland, 1975; Krause et al., 2002), including speaker size in particular

- Difficulties arise when the discrimination involves adult speakers of the same sex
  - Consistent, but not necessarily accurate– why?

Voice Cues to Body Size

Traditionally, two features of the voice have been considered as potential cues to speaker size:

fundamental frequency (F_0) and formant dispersion

determined by the length of the vocal folds in the larynx

determined by the length of the vocal tract

Among adults, there seems to be no consistent relationship between voice pitch and body size within either men or women
(e.g., Collins, 2000; van Dommelen & Moxness, 1995; Rendall et al., 2005)

- Larynx is free to expand in multiple directions

On the other hand:
The length of the vocal tract is more constrained by the anatomical features that comprise it

Voice Cues to Body Size

Fitch (1994, 2000): listeners should be sensitive to voice formants-- a reliable/efficient way of assessing the competitiveness of others

- Evidence of correlations between formants (+ their perceptual salience), VTL, and overall body size in a few different species

  → Does this also hold true for humans?

In humans, the results are more ambiguous

- Modest correlation between vocal formants and height for men, mixed results for women
- Few attempts to assess how well listeners can gauge speaker body size from voice cues
  - Weak or no ability, yet remarkable consistency across raters
    - Speakers with low F_0 rated as large, and those with high F_0 rated as small

The Aim of the Studies

Given that…

discriminating subtle differences in formant frequencies is the basis of vowel and phoneme identification (Stevens, 1998),

why are we not more accurate in our everyday impressions of adult speaker size?

Possible explanations:

- Many of the size differences between adult speakers do not exceed our threshold for formant discrimination (Ives et al., 2005; Smith et al., 2005)
- We are more influenced by (not very reliable) F_0 cues

Experiment 1

Aim: to test listeners' ability to judge the relative size of two speakers based on short stretches of naturalistic speech

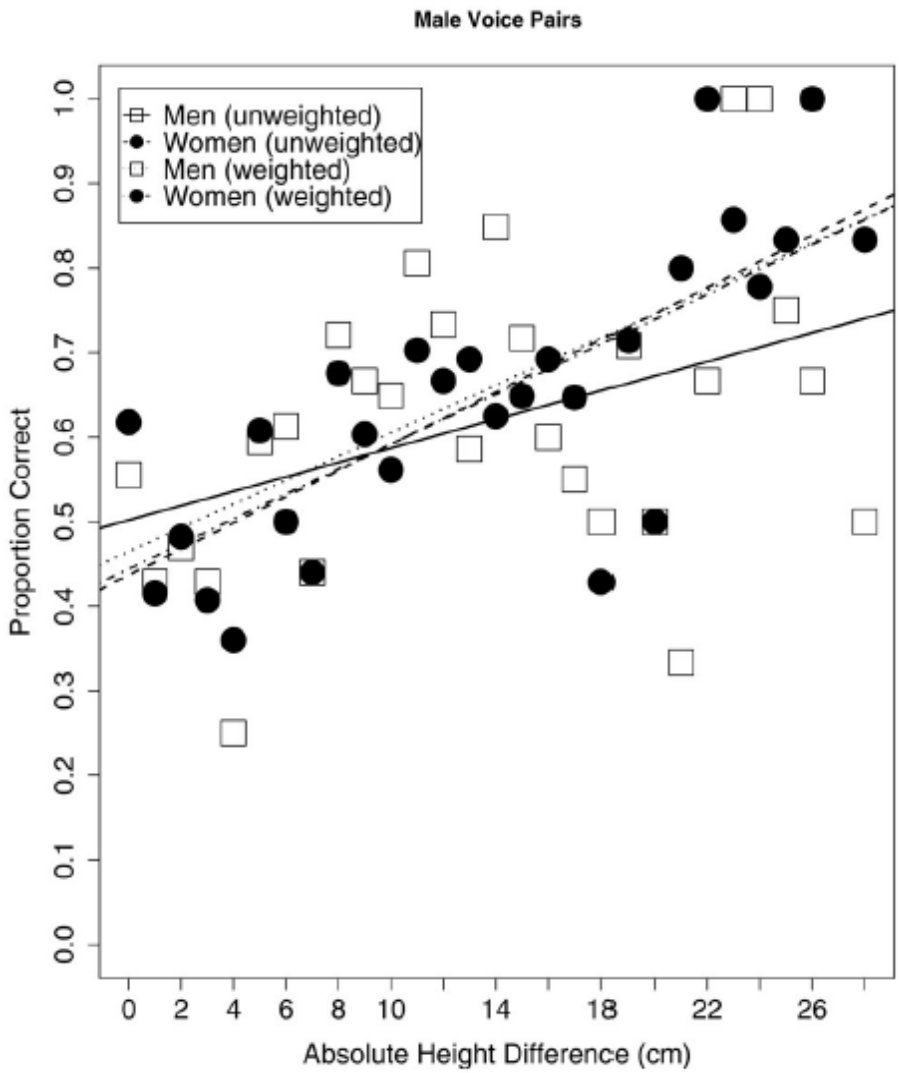
Participants and Stimuli

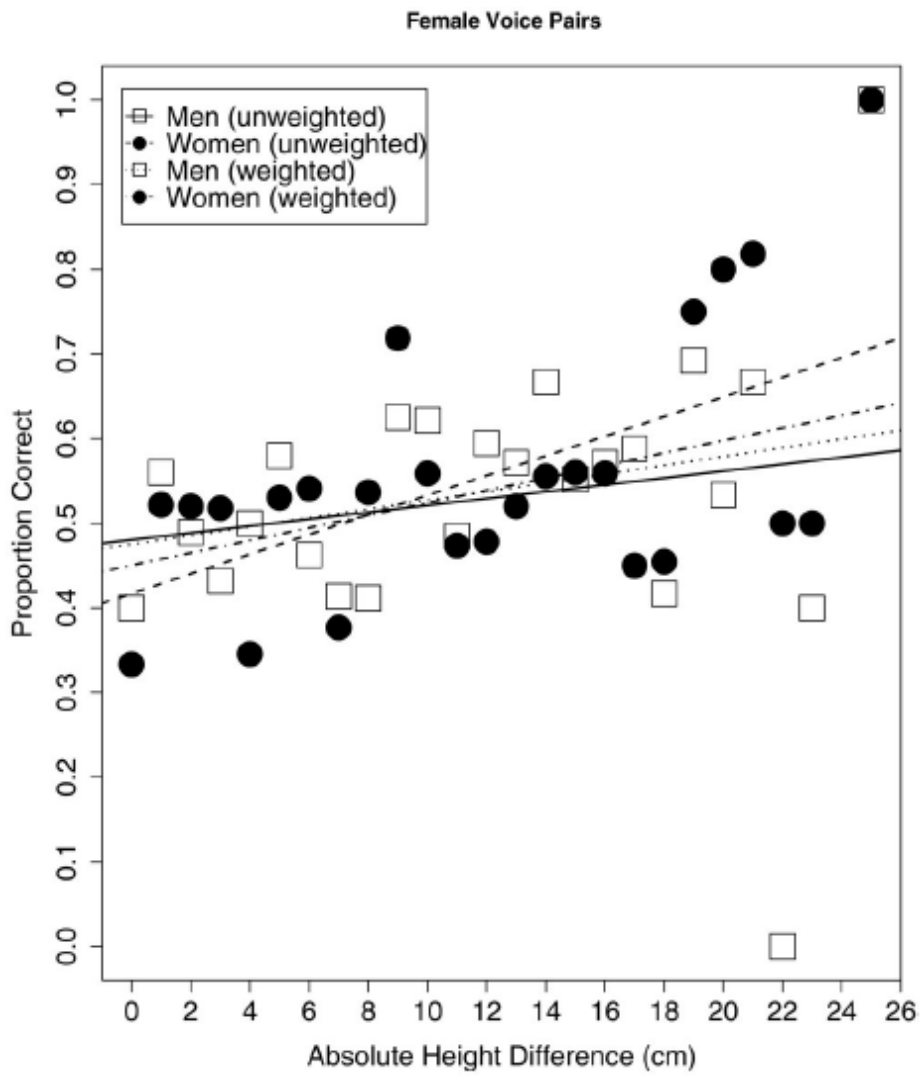
- Listeners: 41 men, 43 women
- Stimuli: 96 phrases from 25 men and 92 phrases from 24 women

Procedure

- Listeners presented with randomly-selected stimuli pairs (two different speakers producing the same phrase)
- Asked which was the larger of the two speakers
  - Two-alternative, forced-choice (2AFC) judgments
- Height used as the index of body size

Experiment 1: Results





Experiment 1: Discussion

The results show that size cues are available and salient in human voices

Asymmetric outcome for listener performance: judgments of female voices were much less reliable

- Consistent with Rendall et al., 2005: acoustic analyses revealed cues to speaker height (contained in formant frequencies) in the voices of males, but not females

Relative success of listeners in this experiment in discriminating speaker size

- Previous work: longer stretches of speech, unpaired testing, open-ended estimates of size
  - More naturalistic, but can complicate the task
    - Dynamic F_0 and formant deviation

Experiment 2

Single-word comparisons of tall versus short male speakers

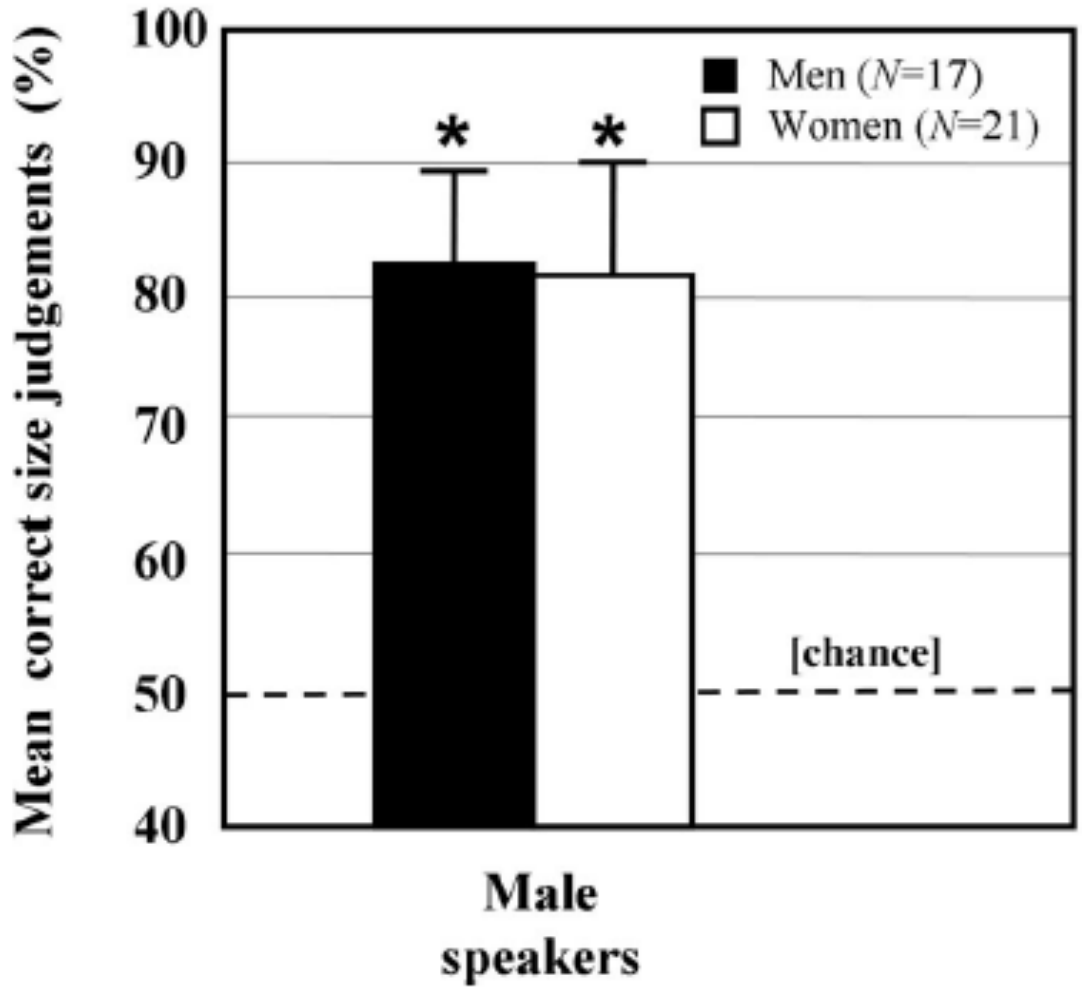
- Reduce the speech material to be compared
- Speakers from the two extremes of the height distribution

→ Listener performance is expected to improve

Method

- Participants: 17 male and 21 female listeners
- Identical materials and procedure to those of Experiment 1, except that the stimuli were limited to single words produced by male speakers who represented the top and bottom 15% of the height distribution for males
  - 6 “short” speakers (≤ 176 cm)
  - 7 “tall” speakers (≥ 190 cm)
  - 25 stimuli pairs in random order

Experiment 2: Results



Experiment 2: Results

Mean height difference between speakers: 19.64 cm (SD = 4.29 cm; range = 14-27 cm)

As expected, listeners performed substantially better in this experiment

- Mean correct responses: 20.45 of 25 trials (81.8%)
- No significant difference between male and female listeners

Formant differences between speakers strongly predicted their height differences, $F(1, 25) = 69.56$

Also a significant relationship between F_0 differences and height differences, $F(1, 25) = 8.64$

→ For some speaker pairs, the listeners could have been using either voice feature

Experiment 3

The role of F_0 versus formants in speaker size assessments

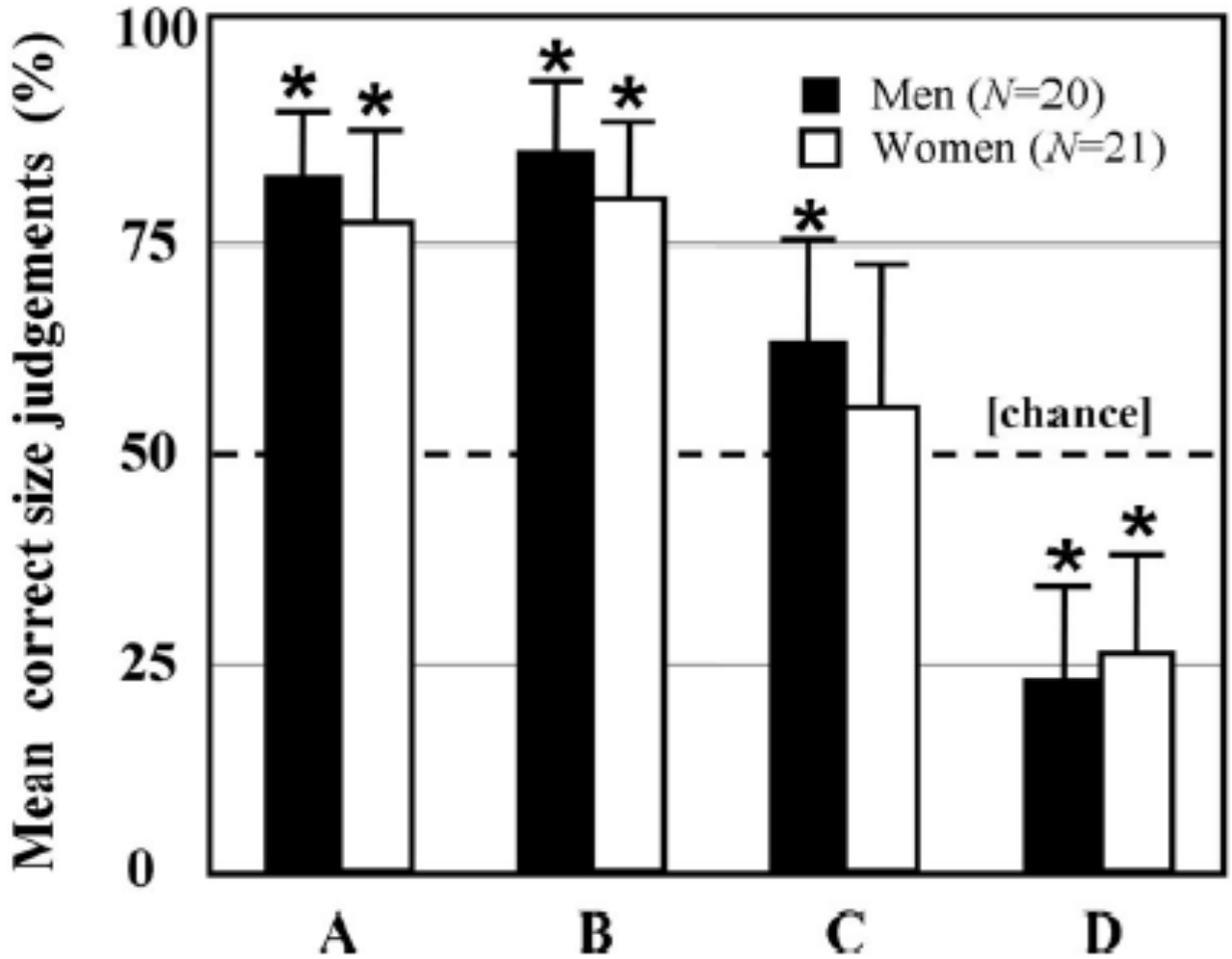
Method

- Participants: 20 male and 21 female listeners
- Four conditions

Neutralize the influence of F_0

- Condition A
  - Matched speakers on F_0
  - If formant cues are salient to listeners, performance in this condition should significantly exceed chance.
- Condition B
  - Flattened the F_0-contour of each word in Praat
- Condition C
  - In half of the pairs from B, the F_0 of the shorter speaker was shifted down by 20 Hz. For the other half, the F_0 of the taller speaker was shifted up by 20 Hz
  - Performance should decline
- Condition D
  - In half of the pairs from B, the formant frequencies of the shorter speaker were shifted below those of the taller male (and reverse modification for the other half)
  - Performance should decline

Experiment 3: Results



→ Listeners are attending to both F0 and formant frequencies

General Discussion

Humans are attuned to voice formant cues to size within classes (i.e., within men)

- Discrimination exceeded chance only when the height difference between speakers was 10 cm or more
  - Consistent with Smith et al. (2005) and Ives et al. (2005): discrimination thresholds for differences in VTL of 4%-8%, height differences of 7-14 cm

Sensitivity to formants could be perturbed by conflicting F_0 information

- F_0 exerts a perceptual pull on size impressions
- If F_0 cues are inherently unreliable, why are our impressions of size often biased by them?

Three possibilities

1. Listeners are overgeneralizing the correlation between F_0 and body size across age and sex classes (Collins, 2000)

2. Applying broader environmental sound-size association to speech
   - Association between large objects/events and low-frequency sounds
   - Sound symbolism

3. Attributing other behavioral traits reliably correlated with F_0
   - Testosterone, body shape variation
     - Evans et al. (2006): Heavier, more broad-shouldered men have lower F_0
     - Low F_0 as a natural marker of status, aggression, dominance?
   - Attractiveness of male voices with low F_0 (Collins, 2000; Feinberg et al., 2005; Putz, 2005)

Thank you!

Questions?

Comments?

Concerns?