Investigating the perception of relative speaker size using synthetic talkers

Santiago Barreda & Terrance M. Nearey

Introduction

Experiments investigating the perception of speaker size typically use stimuli that vary in terms of f0 or their average formant frequencies

- However, these experiments typically control for phonetic content– e.g., by presenting a single vowel category at a time [1,2,3].

This is done so that the influence of f0 and formant-frequency scaling (or vocal-tract length, VTL) on speaker-size judgments may be investigated.

But how does spectral information influence relative-size judgments when the phonetic content of stimuli is not fixed, so that the formants of stimuli covary not only with VTL, but also with vowel category?

Methods

Participants: 19 listeners from the University of Alberta.

Procedure: Listeners heard vowels in pairs. For each trial they were asked to provide a relative height judgment by using a sliding scale.

- 6 vowel pairs were used (see Figure 1, Figure 2).
- Each vowel was also presented with itself.

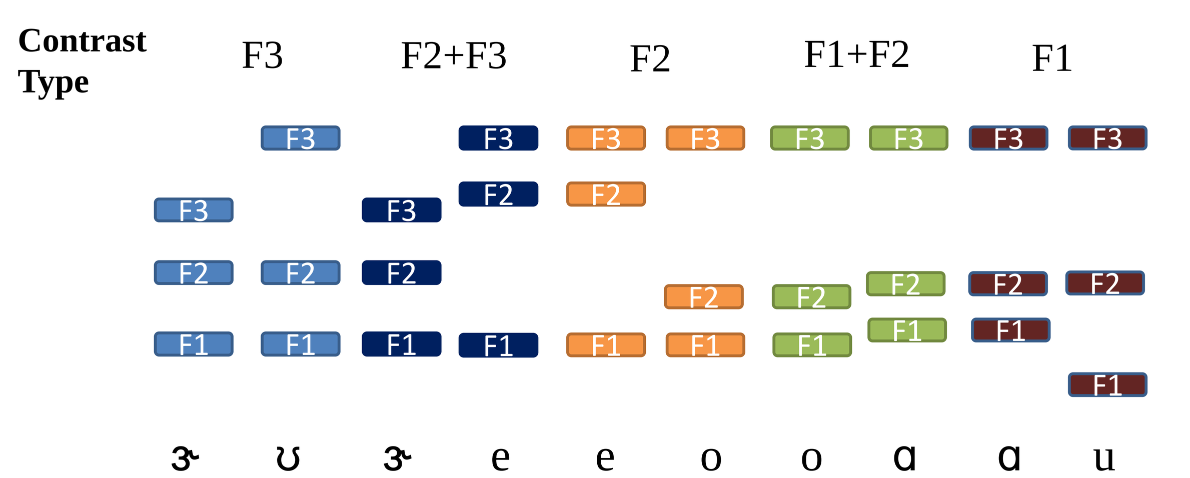


Figure 1 – Schematic of vowel formant patterns for the five vowel pairs used in the experiment. Each pair is indicated using a different color.

Baseline Stimuli: Six vowel categories were used (Table 1). All vowels had steady-state formants, were 200 ms long and had a linearly decreasing f0 from 120 Hz to 110 Hz. Vowels were presented with 300 ms of silence in between.

Vowel	F1	F2	F3	F4
/ ɑ /	700	1150	2500	3400
/ e /	500	2000	2500	3400
/ o /	500	900	2500	3400
/ ɝ /	500	1250	1800	3400
/ ʊ /	500	1250	2500	3400
/ u /	300	1150	2500	3400

Table I. – Formant frequencies of baseline vowel sounds.

Scaling : To simulate VTL differences, vowels were scaled from the baseline voice by increasing and decreasing all baseline voice FFs by fixed percentages.

- Every vowel pair had voices separated by either an 8% or 16% difference in their VTL (or formant-frequency scaling).
- Listeners heard each unique vowel pair (balanced for order) 6 times, resulting in 384 responses per participant.

Rationale for Experimental Design

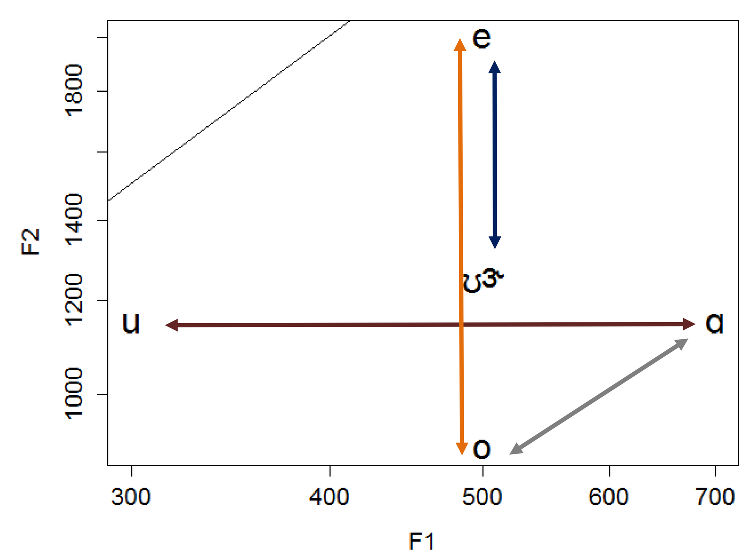
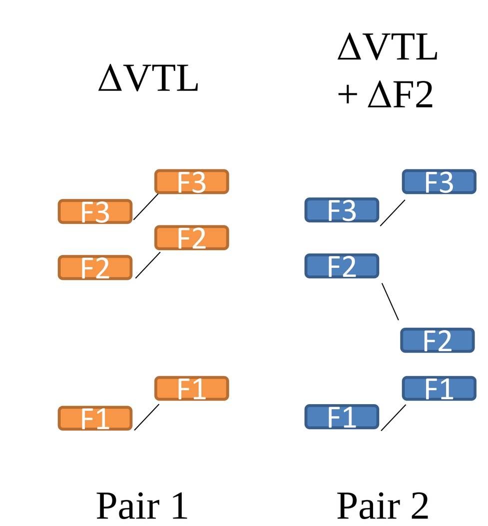


Figure 2 – Experimental vowels at a single VTL level. Lines indicate contrasts made by each vowel pair. The F3 vowel-pair contrast is not indicated in the Figure.

- Vowel pairs differed in either VTL or VTL and 1 or 2 FFs (see in Figures 1, 2).
- If X = the rated height difference between a pair that differs only in VTL.
- And Y = the rated height difference between a pair that differs in both VTL + a formant.
- Then the difference between X and Y may be attributed to that formant.
- This allows for the contribution of individual formants to be investigated.

- Pair 1 differs only in terms of VTL.
- Pair 2 has this difference **plus** an F2 difference.
- The difference in rated height differences between the two pairs can be attributed to the difference in F2.



Results

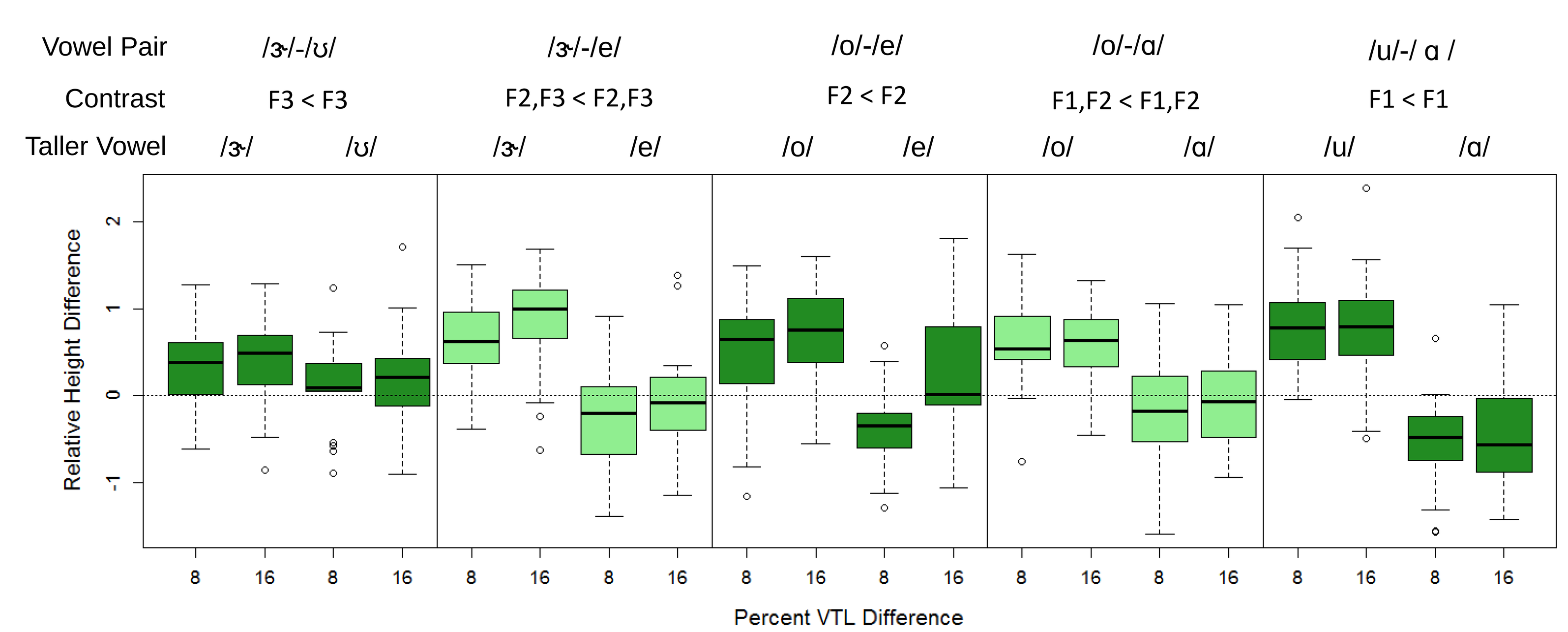


Figure 3. – Distribution of average relative height differences across all listeners, for all two-category vowel pairs. Height differences were standardized within-listener.

- For two-category vowel pairs (Figure 3) and one-category vowel pairs (Figure 4), relative height differences are influenced by both vowel category and VTL difference.

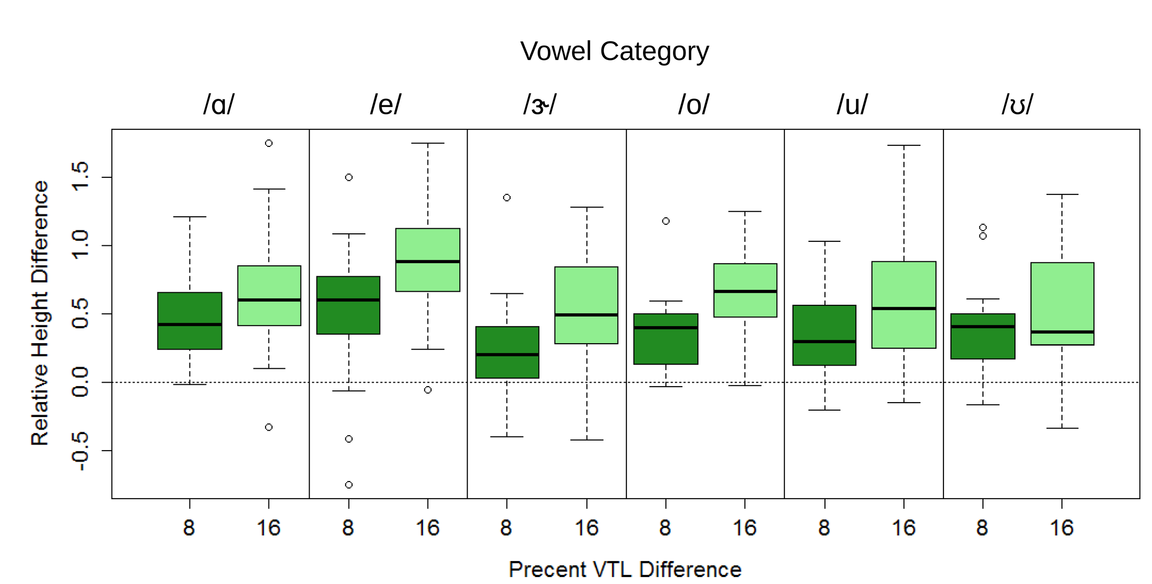


Figure 4. – Distribution of average relative height differences across all listeners, for all one-category vowel pairs. Height differences were standardized within-listener.

- Models were fit individually to each listener's data. Significance testing was carried out across listeners using the fitted coefficients for each listener [4,5].
- For each model, relative height of the first speaker (standardized within listener) was predicted.

- The differences (in log-Hz) between F1, F2 and F3 across the two vowels were used as linear predictors.
- For example, a negative F1 predictor value indicated that the first vowel had a lower F1 than the second.
- VTL differences were coded using dummy variables.
- A Hotelling's T^2 test on the four VTL coefficients revealed that they were jointly significantly different from zero [F(4,15) = 4.47, p = 0.014].

Table II. – Significance tests for the effect of predictors on relative-height judgments.

Effect	Mean	t(18)	p
F1	-0.748	-5.00	< 0.001
F2	-0.489	-3.40	0.003
F3	-0.480	-2.39	0.028
VTL +8%	0.061	-2.95	0.217
VTL -8%	-0.163	0.66	0.019
VTL +16%	0.060	1.28	0.516
VTL -16%	-0.257	-2.56	0.009

- A pooled, fixed-effects model was fit to the data.
- This model was used to get an idea of the predictive power of individual predictors.
- The percent change in variance explained was found between the full model and restricted models excluding one or more effects at a time.

Effect	df	Δ% Variance Explained
F1	1	-24.54
F2	1	-9.55
F3	1	-3.29
VTL	4	-5.38
F1,F2,F3	3	-47.14

Table III. – Change in variance explained associated with the removal of given effects.

Summary and Conclusions

- Relative height judgments are influenced by overall formant-frequency scaling (i.e., VTL).
- Relative height judgments are also strongly influenced by the specific formant-pattern as determined by the vowel category.
    - Large formant-pattern differences (e.g., /u/ vs. /ɑ/ in Figure 3) can overwhelm VTL differences and may result in 'incorrect' judgments.
- No interactions between formant frequency differences and VTL differences were present.

Do listeners make relative height judgments based solely on pattern-corrected VTL estimates?

Probably not, since:

- F1, F2 and F3 account for a good deal of the variance in relative height judgments independently of VTL information.
- Listeners make 'incorrect' assessments when certain vowel categories are paired.

References

[1] Smith, D. R. R. & R. D. Patterson. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. J. Acoust. Soc. A m. 118: 3177-3186.

[2] Lass, N. J. & M. Davis (1976). An investigation of speaker height and weight identification. J. Acoust. Soc. A m. 60: 700-703.

[3] Rendall, D., Vokey, J. R., & Nemeth, C.. (2007). Lifting the Curtain on the Wizard of Oz: Biased Voice-Based Impressions of Speaker Size. J. Exp. Psych: Human Perception and Performance 33: 1208 –1219.

[4] Gumpertz and Pantula. (1989). A Simple Approach to Inference in Random Coefficient Models. The Am. Statistician, 43(4), 203-210.

[5] Lorch, R. F., and Myers, J. L. (1990). "Regression analyses of repeated measures data in cognitive research," J. Exp. Psychol. Learn. Mem. Cogn. 16, 149–157.

Contact Information

Santiago Barreda (sbarreda@ualberta.ca)

Terrance M. Nearey (t.nearey@ualberta.ca)

Department of Linguistics, 4-32 Assiniboia Hall, University of Alberta, Edmonton, Alberta, Canada T6G 2E7