

# Effects of vocal tract growth on gender and vowel identification based on simulated children's vowels



Brad H. Story<sup>1</sup>, Houri K. Vorperian<sup>2</sup>, & Kate Bunton<sup>1</sup>

<sup>1</sup>Department of Speech, Language, and Hearing Sciences, University of Arizona

<sup>2</sup>Waisman Center, University of Wisconsin-Madison



WISCONSIN  
UNIVERSITY OF WISCONSIN-MADISON

## Introduction

The anatomical structures that form the vocal tract exhibit non-uniform growth during the course of development from infancy to adulthood (Vorperian et al., 2009). These growth patterns also indicate sex differences, even at prepubertal stages (Vorperian et al., 2011). It is not well understood how these non-uniform scalings of the vocal tract length affect the formant frequencies, and consequently how they affect both vowel and gender identification.

The purpose of this study was to use a model of the vocal tract to simulate vowels produced by a male talker and female talker at ages ranging from 1 to 20 years, and present them to listeners who were asked to identify both the vowel produced and the gender of the talker.

## Method

### Speech Sample

- Set 1: 160 simulated vowels
- Generated with TubeTalker (Story, 2013) using a nonlinear vocal tract length (VTL) growth algorithm
  - American English vowels /i, æ, a, u/
  - Male vs female growth patterns imposed on male and female vocal tract shapes, respectively
  - Ages 1 to 20 years
  - $f_0$  contour rising and falling with a mean of 280 Hz
  - Vowel duration = 500 ms

### Set 2: 160 simulated vowels

- Identical to Set 1 except the overall VTL was uniformly adjusted.

### Listeners

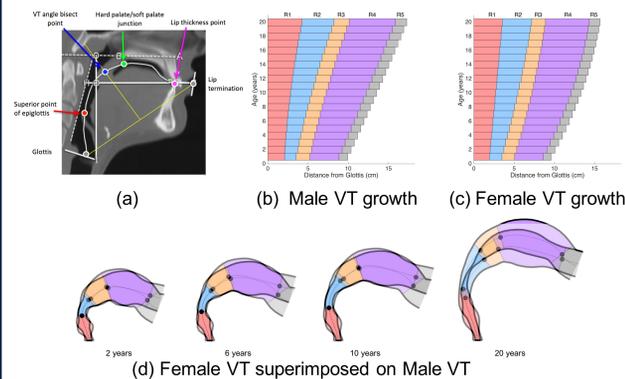
- 16 naïve listeners (13 F, 3 M)
- Mean age = 24.5 years
  - English as native language
  - Passed a hearing screening (0.5, 1, 2, & 4 kHz @ 20 dB)
  - No professional experience with children

### Listening Tasks

- Forced choice vowel id task based on 12 American English vowels using Alvin 3 interface (Hillenbrand, Gayert, & Clark, 2015)
- Identify which of two vowel samples was produced by a boy (paired male vs. female, age matched).

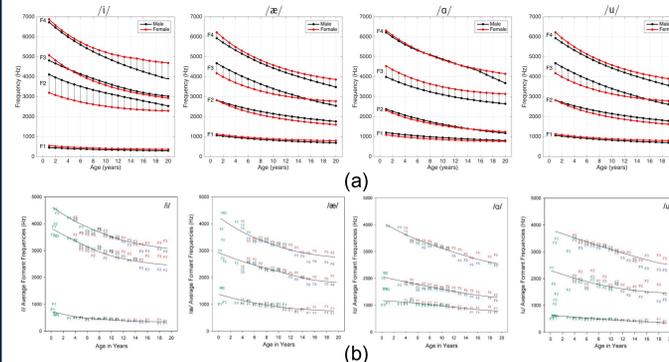
## Model and Simulation

Development of the vocal tract model was based on measurements of sex-specific anatomic landmarks along the vocal tract centerline (Fig. 1a). The measurements define five regions that grow by different amounts over the course of development (Fig. 1b). Vocal tract configurations produced by adjusting the length vector of vocal tract area functions for male/female vowel /a/ (Story, 2005; 2008) are shown at ages 2, 6, 10, and 20 years in Fig. 1c.



**Figure 1:** Sex-specific anatomic measurements and vocal tract growth model for male talker. Color coded regions in (b) & (c) coincide with those in (d).

Using the growth model, a series of VT area functions were generated for the vowels /i, æ, a, u/ at ages 1-20 years (1 year increments) based on previously reported area functions of adult talkers (1 male, 1 female) (Story, 2005; 2008). The calculated resonance frequencies (~formants) of the male and female series for each vowel are shown in Fig. 2a, along with previously reported formant data (Vorperian & Kent, 2007). Two additional series for which the VTL was *uniformly* adjusted were also generated (not shown) as a control set.



**Figure 2:** (a) Calculated resonance frequencies across age compared to (b) reported formant frequencies (from Vorperian & Kent, 2007).

## Results

The results of the vowel identification experiment for both nonuniform and uniform VTL growth across the entire age range are shown in the two confusion matrices below (Tables 1 & 2). In both cases, the /i/ and /u/ targets were identified as these same vowels more than 70% of the time, whereas identification of the /æ/ and /a/ targets was more spread among neighboring vowels.

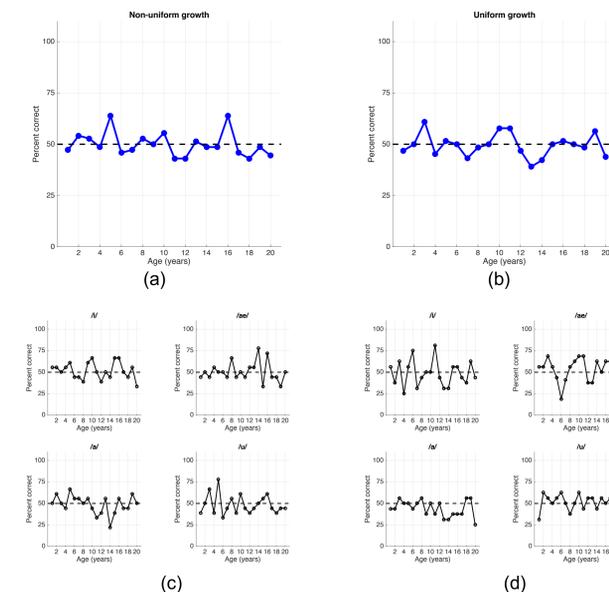
		Identification											
		i	e	æ	a	ɔ	o	u	ʌ	ɜ	ɹ		
Vowel target	i	71	9	2	8	1	0	0	0	0	8	1	0
	e	0	0	0	0	0	0	0	0	0	0	0	0
	æ	2	1	3	26	45	2	5	1	0	1	13	0
	a	0	1	2	3	21	21	27	2	0	1	22	0
	u	0	0	0	0	0	0	0	0	0	0	0	0

**Table 1:** Confusion matrix for *nonuniform* growth vowels

		Identification											
		i	e	æ	a	ɔ	o	u	ʌ	ɜ	ɹ		
Vowel target	i	82	8	1	7	0	0	0	0	0	1	0	0
	e	0	0	0	0	0	0	0	0	0	0	0	0
	æ	1	2	3	26	45	2	8	0	0	0	10	2
	a	0	0	1	2	10	25	51	2	0	1	6	0
	u	0	0	0	0	0	0	0	0	0	0	0	0

**Table 2:** Confusion matrix for *uniform* growth vowels

The gender identification results are shown in Fig. 3 as a function of age for both nonuniform and uniform growth. The top plots contain the mean across the four vowels whereas the lower plots indicate the vowel-specific gender identification. In all cases, the identification oscillates above and below the chance level (50%), exhibiting no particular pattern.



**Figure 3:** Results of gender identification experiment. Shown in all plots is percentage of pairs where the male ("boy") simulation was correctly labeled.

## Discussion

Vowel identification accuracy relative to the targets was fairly high. The confusions of the /æ/ and /a/ vowels with neighboring vowels is expected. In particular, the high rate of identification of the target /a/ as /ɔ/ is likely due to our listener pool from the Southwest US where there is little perceived difference between these vowels.

Even though listeners could determine vowel identity, they were not able to reliably identify the gender of the simulated vowels in either the nonuniform or uniform VTL growth condition.

There are a number of limitations of the study that likely contributed to the inability of listeners to identify gender:

- The vocal tract configuration was static. Articulatory timing may enhance gender specific characteristics but was not implemented in the simulations.
- Scaling of the cross-sectional areas along the VTL was not included in the vocal tract model.
- The voice source was identical for all male and female simulations, and for all ages. This was done deliberately to eliminate voicing cues.

## Conclusions

Although the approach to modeling the growth of the vocal tract produced vowels in accord with published data and were easily identified, it did not, by itself, produce acoustic characteristics that were indicative of gender.

## Acknowledgement

Research supported by NIH R01-DC011275, NIH R01-DC006282, NSF BCS-1145011

## References

- Hillenbrand, J., Gayvert, R. T., and Clark, M. J., (2015). Phonetics exercises using the Alvin experiment-control software. *J. Speech, Lang. Hear. Res.*, 58, 171-184.
- Story, B. H., (2005). Synergistic modes of vocal tract articulation for American English vowels. *J. Acoust. Soc. Am.*, 118(6), 3834-3859.
- Story, B. H., (2008). Comparison of Magnetic Resonance Imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *J. Acoust. Soc. Am.*, 123(1), 327-335.
- Story, B.H., (2013). Phrase-level speech simulation with an airway modulation model of speech production. *Computer Speech and Language*, 27(4), 989-1010.
- Vorperian, H. K., and Kent, R. D., (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *J. Spch. Lang. Hear. Res.*, 50, 1510-1545.
- Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., and Gentry, L. R., (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Acoust. Soc. Am.*, 125(3), 1666-1678.
- Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., and Chung, M. K., (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Spch. Lang. Hear. Res.*, 54, 995-1010.