Normative Lifespan Database for Vowel Formants: Methodological Considerations and Preliminary Results

Houri K. Vorperian, Raymond D. Kent and Diane Austin

Vocal Tract Development Lab - Waisman Center, University of Wisconsin - Madison

Motor Speech Conference
Newport Beach, CA

March 5, 2016

Visit the VTLab online
waisman.wisc.edu/vocal

Introduction

Vowel formant frequencies are used to derive metrics of vowel dispersion, including vowel space area, which presumably reflect articulatory working space. Such acoustic metrics may be sensitive to:

- some types of speech disorder
- developmental and lifespan effects
- sex differences
- variations in speaking style or register

Establishing a sex-specific lifespan normative database of vowel formant frequencies will be instrumental in:

- discovering developmental anatomic-acoustic and articulatory-acoustic relationships
- guiding the interpretation of results for disordered speech in children and adults

Methodological considerations pertinent to the construction of a lifespan normative database of vowel formant-frequency measurements include:

Story-Vorperian-Bunton-MSC2016.pdf

- the recording environment and equipment (microphone/recorder)
- special methodological considerations regarding: 1) the speech stimuli used, 2) acoustic analysis methods and measurements, as well as 3) vowel dispersion metrics.

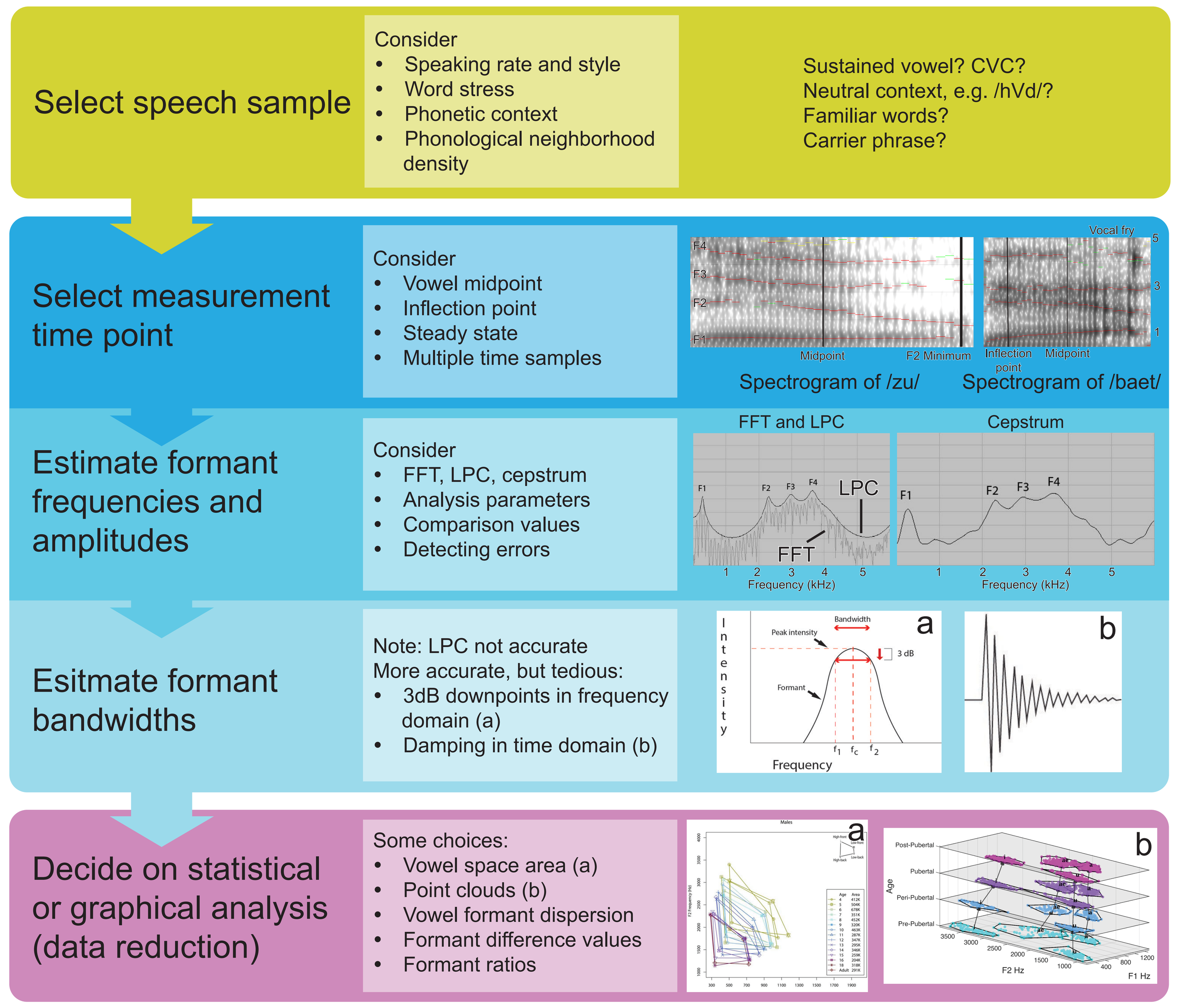

Figure 1. Special methodological considerations for estimating vowel formant frequencies and summarizing the data statistically and/or graphically.

The purpose of this poster is to:

- highlight the importance of these methodological considerations; and to
- present preliminary data on the vocal fundamental frequency (F0) and the first four formant frequencies (F1-F4) for the corner vowels /i u ɑ æ/ for healthy male and female speakers ages 4 to 92 years.

Methods

Participants included 231 healthy individuals (106 male and 128 female participants) between the ages 4 to 92 years.

Ages 4-20 yrs (<20)		
Speaker sex	***n***	**Mean age (SD)**
F	74	11.97 (4.55)
M	64	11.61 (4.15)

Adults (≥20 yrs)				
Speaker Sex	***n* 20-30**	***n* 40-60**	***n* 75-92**	***n* All**
F	21	18	10	49
M	17	9	9	35

Speech stimuli included five monosyllabic words for each of the four corner vowels /i u ɑ æ / as in "eat, hoot, hot and hat". Selected words were:

- familiar to younger participants
- easy to represent as an object or an action
- associated with high phonological neighborhood density, which reportedly maximizes F1/F2 vowel space [6].

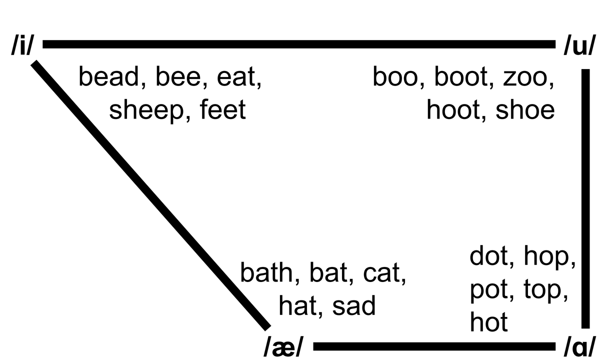

Figure 2. Word stimuli per corner vowel.

Recording protocol entailed recording participants' speech in a quiet room using a Shure-SM48 microphone connected to a Marantz-PMD660 digital audio recorder. The stimuli were presented visually and aurally using a laptop with the TOCS+ Platform program [4] for randomization. Participants were instructed to repeat the words at a normal loudness level. There were 7 tokens per vowel (5 distinct words per vowel, with 2 words repeated).

Acoustic Analysis: Based on findings from previous work that evaluated the accuracy of vowel formant measurements in four acoustic analysis systems [1]; and determined the effect of analysis parameter manipulations on formant measurements in children and adults [3], the following protocol was developed and used:

- Waveforms for each word were segmented using Praat, then the F0 and the F1-F4 frequencies were estimated using a modified version of TF32.
- The spectrogram and waveform of the segmented word were displayed to select the measurement point for formant analysis.
  - Measurement time point was vowel-specific since vowels such as /u/ and /ae/ are often produced with diphthong-like formant shifts, therefore using the typical temporal midpoint of the vowel can yield misleading data.
  - Vowel-specific temporal measurement point used were as follows:
    - /i/ point of highest frequency of F2
    - /u/ point of lowest frequency of F2
    - /ɑ/ point of least separation of F1 and F2 frequencies
    - /æ/ point of most evenly spaced formants (taking care to avoid measurement at a point of decreasing F2-F1 difference which reflects backing of the vowel).
- F1-F4 frequencies were determined with the Fast Fourier Transform (FFT), Linear Prediction Coding (LPC) and cepstrum by inspecting:
  - the spectrogram (with overlaid LPC formant tracks), and the spectral slice (with zoom-in function for greater measurement accuracy)
  - combined displays of the FFT spectrum, LPC spectrum, and cepstrum.
- Vowel-specific analysis challenges included: distinguishing formants that are closely spaced or distinguishing F1 from H1 [first harmonic] in high vowels.
- Analysis parameters manipulated included: dynamic range, DFT points, LPC coefficients, and cepstral coefficients.
- F0 was measured with TF32's pitch tracer at the same temporal point as the formant estimates, except when this point was affected by irregular phonation or the pitch tracer failed. A time-slice FFT was used to identify the first harmonic if the pitch trace was questionable.
- Missing data occurred when it was not possible to make a measurement for a word/vowel. Percent missing by measurement as follows:

	F0	F1	F2	F3	F4
Ages 4-20	0.21%	0.30%	0.10%	2.43%	6.00%
Adults	0.09%	0.04%	0.00%	1.28%	4.85%

Results

Summary plots of the lifespan raw data for male and female speakers are displayed in two panels: <20 (left) and >20 (right) for the fundamental frequency (F0) and formant frequencies F1-F4 for all four vowels in male (blue) and female (red) speakers. Outliers, defined as measurements that exceeded 2.576 SD from their vowel X sex X age group mean, were removed resulting in approximately 1% of measurements being excluded from modeling/analysis efforts.

- **Ages <20 (4-20 yrs):** For each measurement (F0-F4) and for each vowel and sex, a local regression (loess) on age was fit to the word-level data.
- **Ages ≥ 20 (20-92 yrs):** Plots show the mean of the three adult age cohorts (20-30, 40-60 & 75-92 years) connected. Age group comparisons were made separately for males and females using mixed effects models, with fixed effects for sex, age group, and sex X age group interaction, and a random effect for subject. The p-values for the age group comparisons were adjusted using the Tukey-Kramer method.

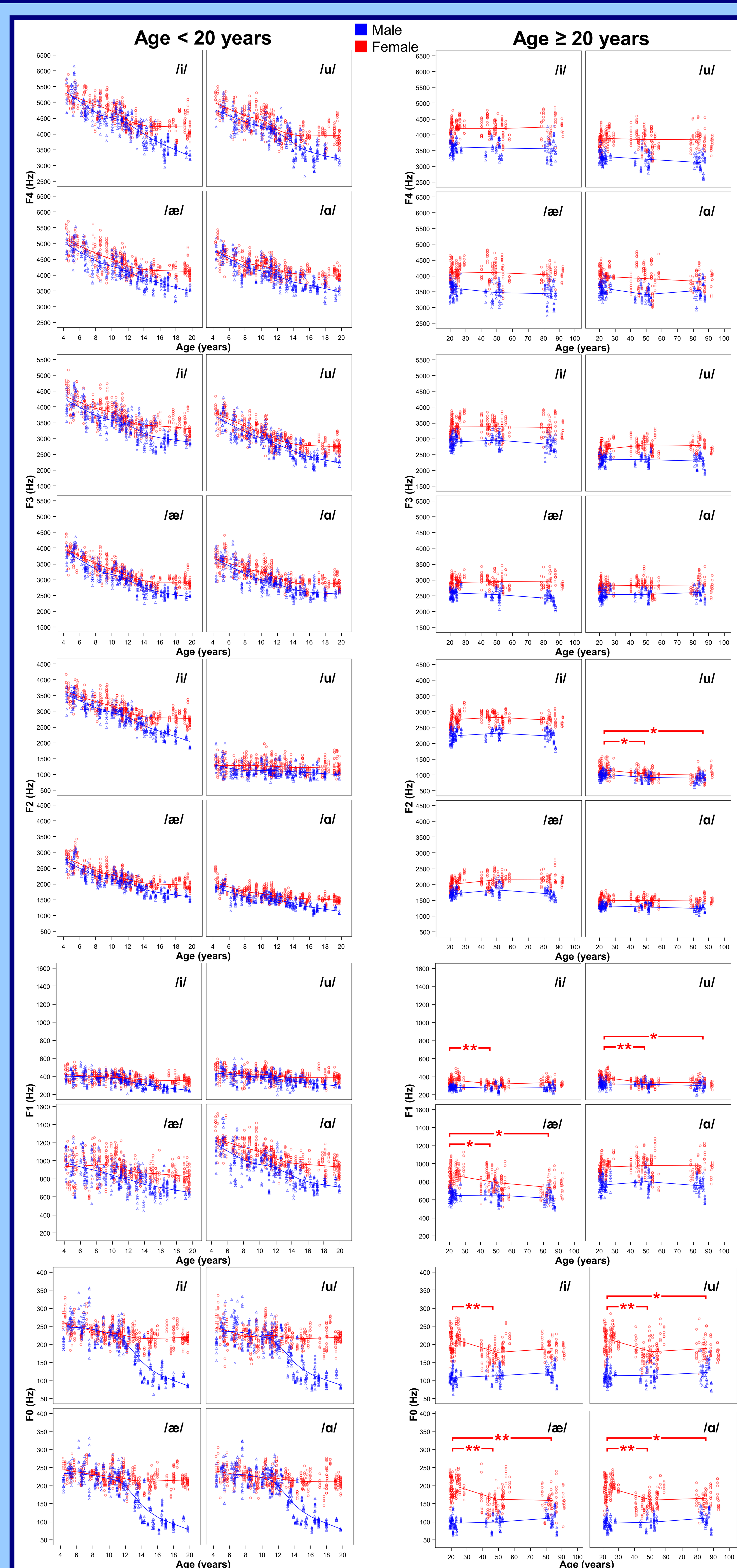

Figure 3. Frequency as a function of age and sex.

Formant and vowel specific developmental trends/observations:

F4
- F4 has a steep developmental gradient for all vowels that tapers off in females during puberty.
- No changes in F4 in adult age cohorts (20-30, 40-60 & 75-92 years).
- Sexual dimorphism evident for all vowels by age 4.

F3
- F3 has a developmental gradient that tapers off in females during puberty.
- No changes in F3 in adult age cohorts.
- Sexual dimorphism evident for all vowels by age 4.

F2
- F2 has developmental gradient for front vowels /i æ/ but not for back vowels particularly /u/.
- F2 decreases significantly for the vowel /u/ in female adult age cohorts.
- Sexual dimorphism evident for all vowels by age 4.

F1
- F1 has a developmental gradient for low vowels /æ ɑ/ only.
- F1 decreases significantly for all vowels except /ɑ/ in female adult age cohorts.
- Sexual dimorphism becomes more evident for the high vowels /i u/ after age 10.

F0
- F0 sexual dimorphism evident after age 10.
- F0 continues to decrease with aging in females but not males.

Discussion – Conclusions

The following conclusions can be drawn from the analyses completed to date:

1. The collection of data with a standardized speech sample and carefully specified methods of analysis may help to reduce extraneous sources of variability so that sex and age differences can be identified more confidently.
2. The pattern of formant frequency change over the developmental period of 4 to 20 years varies with vowel, formant, and speaker sex.
   - The steepest developmental gradients for all 4 vowels were observed for the higher formants F3 and F4. Developmental changes were most prominent in F1 for the low vowels and in F2 for the front vowels.
   - These developmentally sensitive formants are of particular interest in the study of anatomic-acoustic relationships across the lifespan and potentially for the study of speech development in individuals with anatomic anomalies (e.g., Down syndrome).
3. In adults, as age increases, F0 and F1 continue to decrease in female speakers but not male speakers with the largest differences between the young adult and middle adult age groups, and the young adult and older adult age groups.
4. Sex differences in formant frequencies are not uniform across vowels or formants. The earliest indications of sexual dimorphism are observed for the low vowels. Sex differences in formant frequencies are definitely present by age 6 years for all vowels but appear earlier for some vowels.
5. F1 and F2 of vowel /u/ showed little influence of speaker age or sex, and the F2 value was notably lower than in some other published studies[5]. This is likely due to the temporal measurement point used in this study.

Future Direction

1. Examine inter- and intra-subject variability in vowel specific formant frequencies to assess developmental changes in variability.
2. Obtain estimates of formant bandwidth to construct a normative database of bandwidth values.
3. Pursue data reduction using computation of various dispersion metrics to address questions such as vowel-specific vs all-vowel developmental changes [2, 7].
4. Expand vowel formant normative database to include ages 2 to 4 years first, then birth to two years.

References

(1) Burris, C., Vorperian, H.K., Fourakis, M., Kent, R.D., & Bolt, D.M. (2014). Quantitative and descriptive comparison of four acoustic analysis systems: vowel measurements. *J. Speech Lang. Hear. Res.*, 57, 26-45.

(2) Coen, M.H., Vorperian, H.K., & Kent, R.D. (2015). High fidelity analysis of vowel acoustic space. *J. Acoust Soc. Am.*, 137, 2305; 2305 (2015); http://dx.doi.org/10.1121/1.4920418.

(3) Derdemezis, E., Vorperian, H.K., Kent, R.D., Fourakis, M, Reinicke E.L., & Bolt, D. (2015). Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. *Am. J. Speech Lang. Pathol*. doi: 10.1044/2015_AJSLP-15-0020. [Epub ahead of print].

(4) Hodge, M., Daniels, J., & Gotzke, C. L. (2009). TOCS+ Intelligibility Measures (Version 5.3) [computer software]. Edmonton, Canada: University of Alberta.

(5) Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust Soc. Am.,* 105, 1455-1468.

(6) Munson, B., & Solomon, N.P. (2004). The effect of phonological neighborhood density on vowel articulation. *J. Speech Lang. Hear. Res*., 47, 1048-1058.

(7) Vorperian, H.K., & Kent, R.D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *J. Speech Lang. Hear. Res.,* 50, 1510-1545.

Acknowledgement

We thank the following individuals for stimuli development, data collection, and analysis (listed alphabetically): Carlyn Burris, Allison Carolan, Ekaterini Derdemezis, Erin Douglas, Julie Eichhorn, Andrea Kettler, Sara Kutzweil, Katie Lester, Jennifer Lewandowski, Erin Nelson, Allison Petska, Daniel Reilly, Emily Reinicke, Elaine Romenesko, Katelyn Tillman, and Alyssa Wild. We also thank Ben Doherty and Madison Wagner for assistance with poster preparation.

Research supported by NIH-NIDCD R01DC006282 for the project titled MRI & CT studies of the developing vocal tract.