Supplemental Material S1. Detailed description of the VLAM synthesis.

Speech Synthesis

Speech stimuli that conform to the articulatory and acoustic properties of an adult female and infant vocal mechanism (vocal tract geometry and vocal fold vibration) were synthesized using the Variable Linear Articulatory Model (VLAM). VLAM simulates speech that conforms to the vocal production mechanism (vocal tract geometry and vocal fold vibration) for speakers across a broad age range from infancy to adulthood. In doing so, it integrates human vocal tract growth data currently available (Goldstein, 1980) into a previous model already existing for the adult vocal tract (Maeda, 1979, 1990). The latter is based on a statistical analysis of 519 mid-sagittal cineradiographic images of a French speaker uttering ten sentences (Bothorel et al., 1986). A principal components analysis, guided by knowledge of the physiology of the articulators, revealed that seven articulatory parameters (P_i, $i \in \{1\ldots7\}$) could account for 88% of the variance of the tongue contours (Boë et al., 1995). The parameters included labial protrusion, labial aperture, tongue tip position, tongue body position, tongue dorsum position, jaw height, and larynx height. Each parameter is adjustable at a value in the range of ±3.5 standard deviations around the mean values for this articulator in the cineradiographic images. VLAM integrates non-uniform vocal tract growth, in the longitudinal dimension, by two scaling factors: one for the oral cavity and another for the pharyngeal cavity, the zone in-between being interpolated. The values of the factors, from 0.3 to 1.2, were calibrated year-by-year and month-by-month based on Goldstein's (1980) length data. The anatomical measurements of vocal tracts generated by the model are consistent with MRI data, and the acoustic targets are in the range of the mean values of ±1 standard error reported for vowels from 3-year-old to adult speakers in Hillenbrand (1995). The target scaled F_0 values follow those given by Beck (1996). VLAM has also been compared to real infant vocalizations (Serkhane, Schwartz, Boë, Davis, & Matyear, 2007), and it generates ecologically-valid acoustic specifications of vowels.

When synthesizing isolated vowel stimuli, VLAM first generates a mid-sagittal contour of the vocal tract. The two-dimensional mid-sagittal contour is then mapped into a three-dimensional area function using a transfer function (Badin & Fant, 1984). The poles of the transfer function are excited through a five-formant cascade synthesis system (Feng, 1983), by a pulse train generated by a source according to the LF model (Fant et al., 1985). The source parameters (glottal symmetry quotient, and open quotient) are equal to 0.8, and 0.7, respectively, and remain unchanged for all the growth stages. The resulting signal is sampled at 22,050 Hz. The f_0 contours and intensity envelopes are extracted from natural productions of isolated vowels uttered by an adult male speaker.

In the present experiments, VLAM was set to generate five-formant vowels with the vocal tract parameters for adult female and infant speakers. The acoustic targets for the vowels of each speaker type were introduced by manipulating certain anatomical dimensions of the respective vocal tracts. The longitudinal dimension of the vocal tract was modified by varying the ratio between the oral and pharyngeal cavities. The following vocal tract lengths were implemented:

7.70 cm (infant) and 15.36 cm (adult female). Tongue length was calculated to be proportional to palate length. Vocal tract shape was determined using data from Goldstein (1980). Unless specified, VLAM assumes mature motor control abilities to shape the vocal tract in the infant case in order to disentangle the roles of vocal tract anatomy and motor control in the speech output. Different talkers for each speaker type were then generated by manipulating fundamental frequency values (see main text for explanation). Formant values for all of the vowel tokens across the different talker types and audio files of each stimulus are shown in Table S1 below.

Acoustic Description of the Speech Stimuli

Speaker	F_1	F_2	F_3	F_4	F_5	B_1	B_2	B_3	B_4	B_5
Infant	459	4220	5106	6412	8500	61	131	188	468	130
Adult female	313	2628	3115	3704	5222	57	33	151	192	115

Table S1: Mean center frequencies (Hz) and bandwidths (Hz) of the five-formant (F_1, F_2, F_3, F_4, F_5) synthesized /i/ vowels for the adult female and infant speakers simulated for this study.

References

Beck, J.M. (1996). Organic variation of the vocal apparatus. In W. J. Hardcastle & J. Laver (Eds.), *Handbook of Phonetic Sciences*, Cambridge, England: Blackwell, pp. 256-297.

Boe, L-J., Schwartz, J-L., & Valle, N. (1995). The prediction of vowel systems: Perceptual contrast and stability.In: Keller, E. (ed.) *Fundamentals of speech synthesis and speech recognition*. John Wiley, London, 185-213.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P. (1986). *Cinéradiographie des Voyelles et Consonnes du Français* [Cineradiographic study of French vowels and consonants], Institut de Phonétique de Strasbourg, Strasbourg, France.

Fant, G., Liljencrants, J., & Lin, Q. (1985). A four- parameter model of glottal flow. Royal Institute of Technology, *Speech Transmission Laboratory – Quarterly Progress and Status Report*, 4, 1-13.

Feng, G. (1983). Vers unc synthese par la method de poles et des zeros [Toward a synthesis method using poles and zeros]. *Proceedings of the Journees d'Etude sur la Parole*, Groupe Francophone de la Communication Parlee, 155-157.

Goldstein, U.G. (1980). *An articulatory model for the vocal tract of growing children*. Thesis of Doctor of Science, MIT, Cambridge, MA.

Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *Journal of* *the Acoustical Society of America,* 65, S22.

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W.L. Hardcastle & A. Marshal (Eds.), *Speech Production and Speech Modeling* (131-149). Dodrecht, The Netherlands: Kluwer Academic.

Ménard, L., Schwartz, J-L. & Boe, L.-J. (2004). The role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research*, 47, 1059-1080.