

MODELLING THE GROWTH OF THE VOCAL TRACT VOWEL SPACES OF NEWLY-BORN INFANTS AND ADULTS CONSEQUENCES FOR ONTOGENESIS AND PHYLOGENESIS

Louis-Jean Boë

ICP, CNRS INPG/Université Stendhal, Grenoble, France

ABSTRACT

If newly-born infants had the same sensorimotor control capabilities as adults, would their vocal tracts enable them to produce the same range of [i a u] vowel contrasts? Has increasing the volume of the pharynx by lowering the larynx been a necessary evolutionary phase for humans to produce speech? Using a new articulatory model for simulation of vocal tract growth, we calculated the shift and size of the vowel space with age. Our results show that: The maximal vowel space of newborn infants is potentially (at least) the same as adults; There is no reason to think that larynx lowering and increase in pharynx size have been guided by evolution towards speech. Our simulations allow the main flaws of Lieberman's computations to be pinpointed and invalidate his well known thesis presented in 1971 and defended even now: a low larynx and a large pharynx considered as the *anatomical basis* for speech.

1. ANATOMICAL GROWTH DATA

Systematic measurements of the vocal tract from birth to adulthood do not exist at present. However, it is possible to take advantage of cranio-facial measures established at different ages which have been published in anatomy, radiology, and paediatrics. The evolution of the dimensions of the head (osteological structure) and the hyoid bone position (associated, to a certain extent, with the position of the larynx) permit the inference of broad tendencies in the development of the vocal tract. The doctoral thesis of Ursula Goldstein defended in 1980 [8] provides a veritable mine of information: an inventory of data corresponding to 14 distances and 3 angular measurements, established in relation to anatomical reference points and lines, for ages ranging from a few months to 20 years. All of these data can be closely fitted by (double) sigmoidal curves which characterise the general skeletal and muscular growth [2]. Here we summarise and draw attention to the points that are essential in understanding the phenomenon of vocal tract.

At birth, the heads of infants are approximately hemispherical in shape. Increases in the volume and shape of the skull and of the size of the inferior maxilla modify the relative proportions of horizontal and vertical dimensions. The process does not therefore involve a simple uniform scaling, but rather an anamorphosis in which the vertical dimension is emphasised (Figure 1).

For the vocal tract, this phenomenon is further accentuated by lowering of the larynx (inferred, in radiographs, from the position of the hyoid bone). The ratio between the horizontal dimension H (from the articulation of Bjork to the anterior nasal spine, including the thickness of the tissue covering the anterior surface of the superior maxilla beneath the nasal spine) and vertical dimension V (between the line connecting

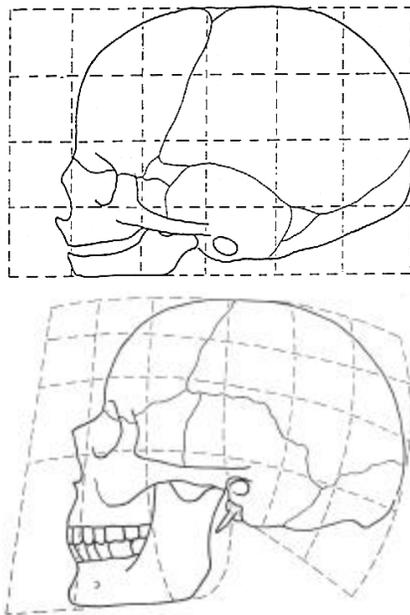


Figure 1. Newborn and adult: anamorphosis of the skull.

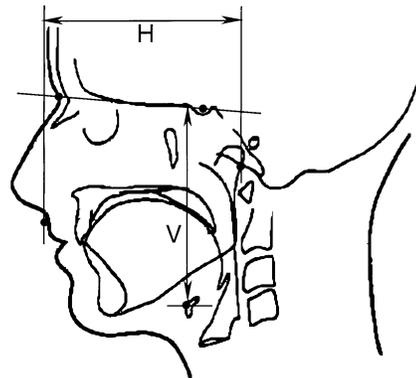


Figure 2. Horizontal (H) and vertical (V) dimension allowing characterization of vocal tract growth.

the sella turcica to the nasion and hyoid bone (Figure 2) varies from 1.6 to 3.0 between birth and adulthood respectively (for male subjects) [10]. Within the vocal tract, the growth of the pharynx is therefore approximately twice as large as that of the front cavity.

Regarding differences between sexes, the data presented by Goldstein [8] demonstrate that there are few differences before the age of 12 years. The growth of the female vocal tract is practically complete towards 15 years, whereas that of the male vocal tract has not yet finished: the male hyoid bone continues to descend, further increasing the length of the pharynx. One can therefore put forward the hypothesis that women would have the same pharynx size as men, if their period of growth were prolonged to span the same length of time. The differences observed between infants, women, and men can thus be considered on the whole to be the consequences of the same basic growth phenomenon.

2. MODELLING THE VOCAL TRACT GROWTH

For adult speakers (male and female), numerous articulatory models of the vocal tract have been available for the past twenty years, established from cineradiographic data and derived from a statistical analysis guided by knowledge of the physiology of the articulators (cf. for example; [7, 15, 23]). These models have the advantage that they intrinsically take into account certain articulatory production constraints: the control parameters are directly interpretable in terms of the degrees of freedom of the articulators (protrusion and labial aperture; movement of the tongue body, dorsum, and tip; larynx height). However, models that possess these characteristics while also simulating growth from birth to adulthood are rare. Here we have used the VLAM growth model (Variable Linear Articulatory Model), developed by Maeda in 1994 [4] which integrates knowledge acquired from previous models with the growth data that are currently available. The growth process is introduced by modifying the longitudinal dimension of the vocal tract according to two scaling factors: one for the anterior part of the vocal tract and the other for the pharynx, interpolating the zone in-between:

$$\begin{aligned} \text{pharynx_scale} &= k(1.1 - 0.30) + 0.30 \\ \text{mouth_scale} &= k(1.0 - 0.65) + 0.65 \end{aligned}$$

The factor k permits the evolution of the vocal tract shape to be simulated, month by month and year by year: this was calibrated using the data provided by Goldstein [8]. The VLAM model was implemented and tested at ICP in an environment originally developed for the SMIP [3]. It is thus suitable for use in systematic simulation studies as well as for use in phonetics.

2. THE CONCEPT OF A MAXIMAL VOWEL SPACE

The models generate a two-dimensional mid-sagittal section, as well as the corresponding area function (three-dimensional equivalent), from which it is possible to calculate the harmonic response (transfer function), formant frequencies (resonance maxima) [1, 9], and speech signal. This procedure is well-suited to modelling vowel production. If the entire input space of command parameters is explored — while satisfying the conditions necessary for vowel production — one can simulate the maximal F1-F2-F3 acoustic space appearing at the output. All possible oral vowels are thus situated within the limits of this region [5]. This kind of extended generation method allows possibilities for maximal distinctiveness to be described precisely, and permits an optimal choice of prototypical realisations. Such an approach can be shown to be more reliable than one that consists in extrapolating the limits of the vowel space for a particular vocal tract (or model) from three unique examples corresponding to [i a u], which are not guaranteed to be optimal (the flaw of Lieberman and Crelin [14]).

4. THE INFLUENCE OF VOCAL TRACT GROWTH ON THE MAXIMAL VOWEL SPACE

4.1. Detailed global predictions

By using a schematic representation of the vocal tract limited to 2 or 4 tube sections (twin tubes) in conjunction with a lossless acoustic model, Mol [17] established in 1970 the basic general tendencies for this phenomenon. In the case of a simple linear transformation applied to the length L (kL with $k=1$ for an adult and $k<1$ for an infant), the maximal vowel space F1-F2 is scaled homogeneously by a factor of $1/k$. This global result agrees quite well with the well known data [20], established for adults (male and female) and adolescents for the vowels [i a u œ], for example, which define the limits of the American English vowel space (for more details [6, 19]).

Using formant-cavity affiliation relationships, it is possible to predict the consequences of non-linear growth on the F1-F2 and F2-F3 spaces. By adjusting the prototypical area functions for [i a u] estimated from Maeda's articulatory model [15] using 2-tube and 4-tube models, it is possible to determine the following affiliations [16], and to predict the influence of the volume and length of any particular part of the vocal tract (the pharynx, as it happens) on vowel contrasts (Table 1). A linear length transformation with scale factor k results in a scaling by $1/k$ of the formants associated with resonance $n\lambda$, as is also the case for the Helmholtz resonances, on condition that the ratio between the neck area ($A_{\text{neck}} l_{\text{neck}}$) and cavity area (A_{vol}) is held constant.

For the maximal vowel space delimited by [i a u] it can be predicted [4] that:

- a relatively linear transformation of F1 and F2 will occur: F1 [i] and F1 [a] depending on the front cavity; by examining affiliations, F2 [i] and F2 [u] are found also to be affiliated, for infants, with the front cavity;
- a contrast between F2 and F3 will occur, which is less pronounced for infants; these two formants become associated with front and back cavities of approximately equal length (between 3 and 4 cm).

4.2. Simulations with the VLAM model: the maximal vowel space and vowel prototypes

Using VLAM, we generated a set of vowels for a grid of command parameters P_i ($-2 \leq P_i \leq 2$ in steps of 0.5), constraining the intra-oral constriction and lip area to be identical for adult and neonate (constriction area 0.3 cm^2 and lip area 0.15 cm^2). Table 2 presents the limits of this space (in Hz, and in perceptual units, Bark and ERB [18]). It is evident that the perceptuo-acoustic space in the F1-F2 plane is at least as large, if not larger, for newly-born infants as it is for adults (Figure 4), and that in the F2-F3 plane it is clearly reduced (from 3.2 Bark or 4.3 ERB).

From this maximal vowel space, we determined prototypes for the vowels [i a u], taking into account the positions proposed by Goldstein [8]. The VLAM model thus permitted us to calculate (1) the articulatory configurations that a new-born child would have if it were to display the same control capacities as an adult, (2) as well as the configurations for an adult, and for an adult having undergone a linear growth of the vocal tract, (3) with and (4) without length normalisation. The tendencies noted in the previous paragraph can be verified.

We tested the VLAM model using the data given by Patricia Kuhl [11], which were obtained from infants aged 3, 4, and 5 months using an imitation task. The formants for [a] correspond

closely to our predictions, whereas those for [u], and to a lesser extent those for [i], are not attained: children do not yet possess sufficient control for the latter two vowels.

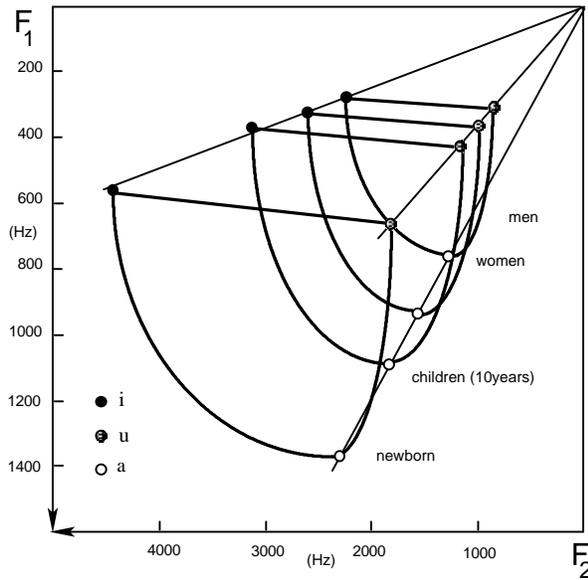


Figure 3. The predicted Maximal Vowel Space, for men women, children and newborn infants (from [17], in accordance with [20]).

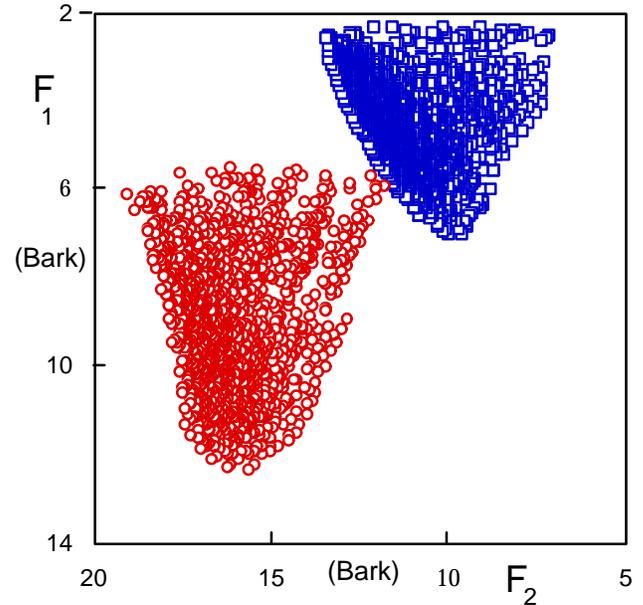


Figure 4. The Maximal Vowel Space generated by the growth model (adult and newborn infant).

	F ₁ max – F ₁ min			F ₂ max – F ₂ min			F ₃ max – F ₃ min		
	(Hz)	Bark.....ERB)		(Hz)	Bark.....ERB)		(Hz)	Bark.....ERB)	
adulte	546	4.7	7.5	1387	6.2	8.1	2018	5.2	6.4
nouveau-né	1302	7.1	9.5	3246	7.2	8.7	1658	2.0	2.1

Table 2. F1-F2-F3 limits for the Maximal Vowel Space established using the VLAM.

	L (cm)	F ₁		F ₂		F ₃		
		(Hz)	(Bark)	(Hz)	(Bark)	(Hz)	(Bark)	
[i]	1. newborn	7.7	628	6.0	4824	18.9	5714	20.1
	2. adult	17.4	246	2.6	2058	13.1	3325	16.3
	3. linear growth	13.7	360	3.7	2902	15.4	3474	16.6
	4. linear growth + norm.	17.4	306	3.2	2293	13.8	2846	15.3
[a]	1. newborn	7.3	1489	11.0	3477	16.7	5430	19.7
	2. adult	17.3	678	6.4	1275	10.0	2217	13.6
	3. linear growth	13.0	913	8.0	1817	12.3	3041	15.7
	4. linear growth + norm.	17.3	720	6.7	1383	10.5	2304	13.8
[u]	1. newborn	8.6	538	5.3	1614	11.5	5356	19.6
	2. adult	19.4	299	3.1	682	6.4	1985	12.8
	3. linear growth	15.1	365	3.7	889	7.8	2913	15.4
	4. linear growth + norm.	19.4	312	3.2	706	6.6	2267	13.7

Table 3. Vowel prototypes for [i a u] established using the VLAM model. Vocal tract length L (in cm), formants F_i (in Hz and Bark).

5. CONSEQUENCES FOR ONTOGENESIS AND PHYLOGENESIS

If the growth of the vocal tract from newly-born infant to adult were linear, the effect would be to shift the maximal vowel space in frequency, while preserving the same perceptual contrasts (regardless of whether these are expressed in Bark or ERB rate): in the F1-F2 plane these contrasts would be enhanced, whereas in the F2-F3 plane (exploited in only 7% of known languages) these would be relatively reduced. The non-proportional growth of the vocal tract due to larynx lowering does not therefore present a decisive advantage for the production of a large maximal vowel space, since the F1-F2 and F2-F3 contrasts are only better balanced perceptually.

If newly-born infants had the same sensorimotor (control) capacities as adults, their vocal tracts would allow them to produce an F1-F2 vowel space as extensive as that of their parents. They would simply need time to acquire and master the relevant control strategies (in particular, the dual control necessary for [u]).

Modelling the growth of the vocal tract enables a unification of our understanding of the phenomena governing inter-individual differences between neonates, babies, adolescents, and male and female adults. It opens up new operational perspectives for normalisation procedures.

These results confirm the modelling work described by Goldstein [8], and are completely consistent with the existing data that have been collected: as far as we know, nobody to date has claimed that adolescents and women — who have shorter pharynges than men — have difficulty in realising vowel contrasts!

Endowed with a small pharyngeal cavity, monkeys exhibit the same vocal tract configuration as newly-born infants, but if they do not produce vowels, it is not due to this resemblance. This is, however, the hypothesis proposed by Lieberman and Crelin in 1971 [13], and republished regularly ever since [12] a hypothesis that is systematically presented as fact in numerous publications, encyclopedias, and works of reference.

“Note that the chimpanzee and newborn human utterances only cover a small portion of the ‘vowel space’. In other words, the chimpanzee and newborn vocal tracts according [our] modelling technique inherently do not appear to be able to produce vowel [a], [u], and [i] [...] It is not a question of the chimpanzee and newborn not being able to produce American-English vowels. They could not produce the vowel range that is necessary for any other languages of Homo Sapiens” Lieberman [13].

No, if monkeys do not talk, according to present evidence this is due to a lack of appropriate cortical equipment (perhaps through lack of differentiation of control between the larynx and articulators), and a lack of sufficient cognitive capabilities. The brain is entirely capable of controlling a vocal instrument with a somewhat longer or shorter pharynx: these differences do not actually change the capacity of contrasting maximally vowels.

ACKNOWLEDGEMENTS

This research was partly carried out within the framework of a project financed by the Agence Rhône Alpes des Sciences Sociales et Humaines *A generator of the sound structures of language* and an international project in Cognitive Science

Robots speak to robots. The model was handled in collaboration with S. Maeda. It benefited, in particular, from the attentive stimulation, advice, and criticism of C. Abry, P. Badin, J.L. Schwartz, I. Maddieson, C. Stoel-Gammon, and M. Vihman. Special thanks to G. Ramsay for translation.

REFERENCES

- [1] Badin, P. and Fant, G. 1984. Notes on vocal tract computations. *STL QPSR* 2-3, 53-108.
- [2] Bock, D., Wainer, H., Petersen, A., Tissen, D., Murray, J. and Roche, A. 1973. A parametrization for individual human growth curves. *Human Biology* 45, 1, 63-80.
- [3] Boë, L.J., Gabioud, B. and Perrier, P. 1995. Speech MAPS Interactive Plant SMIP. *XIIIth Int. Congr. of Phonetic Sciences* 2, 426-429.
- [4] Boë, L.-J., Maeda, S. 1998. Modélisation de la croissance du conduit vocal. Journées d'Études Linguistiques “*La Voyelle dans tous ses états*”. Nantes, 98-105.
- [5] Boë, L.J., Perrier, P., Guérin, B. and Schwartz, J.L. 1989. Maximal Vowel Space. *Eurospeech* 89, 2, 281-284.
- [6] Fant, G. 1975. Non-uniform vowel normalization. *STL-QPSR* 2-3, 1-19.
- [7] Gabioud, B. 1994. Articulatory models in speech synthesis. In Keller E. (ed.), *Fundamentals of Speech Synthesis and Speech Recognition*, 215-230, J. Wiley & sons, Chichester.
- [8] Goldstein, U.G. 1980. *An articulatory model for the vocal tract of the growing children*. Thesis of Doctor of Science, MIT, Cambridge, Mass.
- [9] Jospa, P. 1992. Caractérisation variationnelle des modes de résonance dans le conduit vocal. *Rapport d'Activité de l'Institut de Phonétique de Bruxelles*, 28, 13-30.
- [10] King, E.W. 1952. A Roentgenographic study of pharyngeal growth. *Angle Orthodontist* 22, 23-37.
- [11] Kuhl, P. 1995. Vocal learning in infants: development of perceptual-motor links for speech. *XIIIth Int. Congr. of Phonetic Sciences* 1, 146-149.
- [12] Lieberman, P. 1973. On the evolution of language : A unified view. *Cognition* 2, 1, 59-94.
- [13] Lieberman, P. 1991. *Uniquely Human, The evolution of speech, thought, and selfless behavior*. Harvard Univ. Press, Cambridge, Mass.
- [14] Lieberman, P. and Crelin, E.S. 1971. On the speech of the Neanderthal man. *Linguistic Inquiry* 2, 2, 203-222.
- [15] Maeda, S. 1989. Compensatory Articulation during Speech. In Hardcastle, W.J. and Marchal, A. (eds.) *Speech Production and Modelling*, 131-149. Academic Publishers, Kluwer.
- [16] Maeda, S. and Carré, R. 1995. Modèles de production. In *École Thématique Fondements et perspectives en traitement automatique de la parole*. GDR-PRC Communication Homme-Machine, 51-72.
- [17] Mol, H. 1970. *Fundamental of Phonetics II: Acoustical modes generating the formants of the vowel phonemes*. Mouton, La Haye.
- [18] Moore, B.C.J. and Glasberg, B.R. 1983. Suggested formulae for calculating auditory-filter bandwidth and excitation patterns. *J. Acoust. Soc. Am.* 93, 74,3, 750-753.
- [19] Nordström, P.-E. 1977. Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics* 5, 81-92.
- [20] Peterson, G.E. and Barney, H.L. 1952. Control method used in the study of vowels. *J. Acoust. Soc. Am.* 24, 2, 175-184.
- [21] Vallée, N. 1994. *Systèmes vocaliques: de la typologie aux prédictions*. Thèse de Doct. Sci. Langage. Université Stendhal, Grenoble.
- [22] Vallée, N., Boë L.J. and Payan J. 1995. Vowel Prototypes for UPSID's Phonemes. *XIIIth Int. Congr. of Phonetic Sciences* 1, 424-427.
- [23] Vilain, A. 1997. *Un nouveau modèle articuloire pour la synthèse et le contrôle robotique de la parole: Gentiane*. DEA Sci. du Langage. Univ. Stendhal, Grenoble.