

Journal of Phonetics (to be published)

The potential Neandertal vowel space was as large as that of modern humans

Louis-Jean Boë

Institut de la Communication Parlée, UMR CNRS, Grenoble, France

Jean-Louis Heim

Laboratoire d'Anthropologie, Musée de l'Homme, Paris, France

Kiyoshi Honda

ATR, Kyoto, Japan

and

Shinji Maeda

École Nationale Supérieure des Télécommunications, Paris, France

Since Lieberman & Crelin postulated in 1971 the theory that Neandertals were a speechless species, the potential speech capability of Neandertals has been the subject of hot debate. Lieberman and Crelin claimed that the development of a low laryngeal position was a necessary condition for the realization of a sufficient number of vocalic contrasts, since the potential vowel space was enlarged due to an enlarged pharyngeal cavity. Like newborn infants, Neandertals did not possess this "anatomical basis of speech", and therefore could not speak. Lieberman and Crelin further claimed that this fact may have caused the, otherwise mysterious, extinction of the Neandertal. In this study, we refute the articulatory and acoustic arguments developed by Lieberman and Crelin in their theory. Using a new anthropomorphic articulatory model, we infer that the vowel space of the Neandertal male was no smaller than that of a modern human, and we present vowel simulations to corroborate this hypothesis. Our study is strictly limited to the morphological and acoustic aspects of the vocal tract, and we cannot therefore offer any definitive answer to the question of whether Neandertals spoke or not. However, we do feel safe in claiming that Neandertals were not morphologically handicapped for speech. A low larynx (and large pharynx) cannot be considered to be the "anatomical prerequisites for producing the full range of human speech". There is therefore no reason to believe that the lowering of the larynx and a concomitant increase in pharynx size are necessary evolutionary pre-adaptations for speech.

1. A largely controversial and widespread theory

Lieberman & Crelin (1971), henceforth L&C, followed up by Lieberman (1972, 1973, 1984, 1991, 1994) shocked paleontologists and anthropologists, as well as the speech science community, when they reconstructed the shape of the vocal tract from the fossilized skull of the Neandertal man from *La Chapelle-aux-Saints*, and estimated the extent of his vowel space. The vowel space turned out to be highly reduced with respect to that of a modern human male. They concluded that Neandertals could not have possessed speech, as modern humans do, and that this was probably the cause of their mysterious extinction about 30,000 years ago.

E-mail: boe@icp.inpg.fr, heim@mnhn.fr, maeda@sig.enst.fr, honda@hip.atr.co.jp

According to L&C, in order for speech production to become possible during the course of evolution, it was necessary that the skull base be sufficiently flexible, and that the larynx descend to enlarge the pharyngeal cavity for the vowel space to be large enough to realize the contrasts observed in current human vowel systems (Laitman, 1983; Laitman, Heimbuch & Crelin, 1979).

Vowel spaces are normally delimited by the three extreme vowels, [i], [a] and [u], which are the most common, and are present in a large majority of world's languages (Maddieson, 1986; Vallée, 1994; Schwartz, Boë, Vallée & Abry, 1997a). The development of a proper vocal-tract morphology is a necessary condition for our ancestors to have articulated sounds of speech, with of course, the concomitant developments of appropriate motor control and cognitive functions. L&C advanced the hypothesis that Neandertals did not possess this necessary "anatomical basis", and therefore could not speak. In an acoustic simulation study, they demonstrated that the contrast between the three extreme vowels was highly reduced (resulting in a small vowel space) for a male Neandertal vocal tract as well as for the vocal tracts of a modern newborn human and of a chimpanzee.

In their revolutionary proposal, L&C grouped together a Neandertal, a chimpanzee, and a human infant in the same class, all having a short pharyngeal cavity relative to the oral cavity and thus being incapable of articulate speech. The lowering of the larynx and the increased size of the pharynx have been guided by evolution towards speech.

The way in which L&C reconstructed the vocal tract of Neandertal man, based mainly on anatomical considerations (in particular the position and the arrangement of the hyoid-larynx complex) has been criticized as unrealistic by a number of authors (Falk, 1975; Trinkaus & Shipman, 1993; Schepartz, 1993; Houghton, 1993; McCarthy & D. Lieberman, 1997). Moreover L&C employed a cast of the skull of the man from *La Chapelle-aux-Saints* that was reconstructed by Boule (1913, 1921) with certain errors, notably in the region of the skull base, which became evident after its recent reconstruction by Heim (1986, 1989, 1990).

In spite of numerous criticisms, the thesis proposed by Lieberman and Crelin is systematically presented as fact in numerous publications, encyclopedias, and works of reference. In this paper, we refute the articulatory and acoustic arguments developed by L&C in their theory by:

- Quantifying the vocal tract geometry by a Laryngeal Height Index (LHI) corresponding to the length ratio of the pharyngeal cavity to the oral cavity. Using new biometric data, we try to estimate the laryngeal position for two Neandertals.
- Using a new anthropomorphic articulatory model, we introduce variations of the LHI corresponding to that of a vocal tract of a newborn infant, a child, a female adult and a male adult. We quantify the potential maximum vowel space corresponding to these LHI variations and we compare the results using length normalization.

2. Estimation of the larynx position of fossils

2.1. Qualitative method

The characteristics of the skull base, of the mandible and of the mastoid and styloid processes (on which the muscles suspending the hyoid-larynx complex insert) allow paleontologists to estimate the position of the larynx and hence the dimension of the pharynx (Laitman *et al.*, 1979; Laitman, 1983; Arensburg, Bar-Yosef, Goldberg, Laville, Meignen, Rak, Tchernov., Tillier & Vandermeersch, 1985; Arensburg, Tillier, Vandermeersch, Duday, Schepartz, & Rak, 1989; Heim, 1986, 1989, and 1990; MacCarthy & D. Lieberman, 1997; D. Lieberman &

MacCarthy, 1999). Errors can occur in the reconstruction however, especially when comparative models are absent. For example, we now understand that Boule made various mistakes in the reconstruction of the skull of the Neandertal man from *La Chapelle-aux-Saints*, since at that time Neandertals were hardly known and Boule used a chimpanzee as a model for the reconstruction. Since 1984, Heim (1989) has undertaken the new reconstitution of the skull of the man from *La Chapelle-aux-Saints*. He has shown in the new reconstruction that the position of the head and the skull base does not fundamentally differ from that of a modern human as far as the angulation of the basicranium is concerned (Heim, 1989) and concludes that, in all likelihood, the larynx of the Neandertal was located at the same position as that of modern human populations (Heim, 1986, 1989, 1990). Moreover, after the discovery of a 60,000 year-old Neandertal with the complete hyoid bone intact in a sepulture at the site of Kebara in Israel, Arensburg *et al.* (1989) stated that "There has been little or no change in the visceral skeleton (including the hyoid and inferentially the larynx) during the past 60,000 years of human evolution", which agrees with Heim's conclusion.

2.2. Quantitative method

Recently, Honda & Tiede (1998), henceforth H&T, have shown using magnetic resonance imaging (MRI) on modern human subjects that the vertical larynx position can be predicted from biometric measures of the cranio-mandibular geometry. Their basic idea is that the facial geometry of humans, and more generally that of primates, is related to the position of the larynx. To quantify a global facial morphology, H&T propose the following three distances, palatal distance (PD), laryngeal height (LH) and oral cavity height (OCH). PD is the distance between the anterior nasal spine (ANS) and the posterior nasal wall (PNW). PNW is defined as the intersection point of a standard palatal line (specified by ANP and the posterior nasal spine) and the posterior nasopharyngeal wall. LH is the distance from the arytenoid apex to the palatal line. LH, therefore, represents the pharyngeal cavity length and the vertical position of the larynx. OCH is defined as the distance from the gnathion to the palatal line. Figure 1 illustrates these landmarks and distances which are superimposed on our modified landmarks described later.

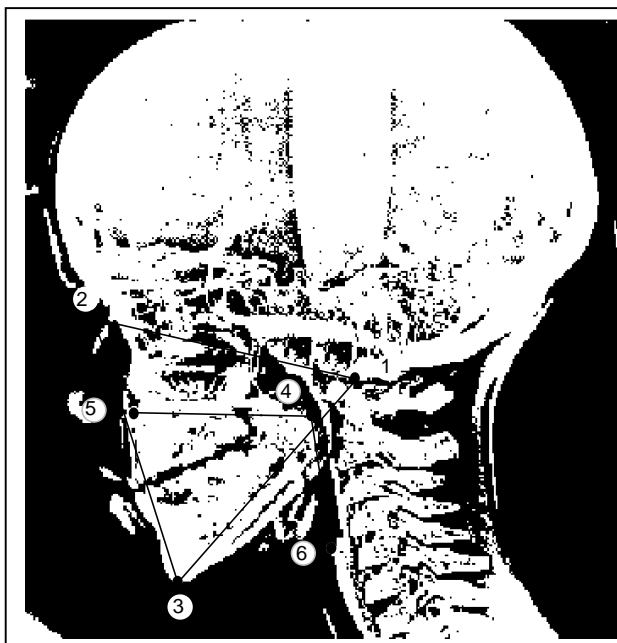


Figure 1. The three landmarks: (1) basion, (2) nasion and (3) gnathion are used in this study to infer the original landmarks (4) the posterior nasal wall, (5) the anterior nasal spine proposed by Honda and Tiede (1998) to characterize a global morphology of the cranium and mandible, which can be used predict (6) the arytenoid apex, i.e., laryngeal position (Xerography courtesy of Denis Autesserre).

H&T have measured these distances on MR images of adult subjects with an age span of 23 to 48 years. Their subjects consisted of 12 Japanese speakers (2 females and 10 males) and 12 non-native North Americans (3 females and 9 males). The results indicate that when LH and OCH are normalized by PD, using a laryngeal height index ($LHI = LH/PD$) and oral cavity index ($OCI = OCH/PD$), the two indices exhibit a high degree of correlation ($r = 0.866$). Using this correlation, we can predict the value of LH from OCI and PD, and thus the position of the arytenoid apex, which is often missing in the fossil records of Neandertals and in anthropological records.

To compare the laryngeal position of Neandertals with that of modern humans, we must have some idea about the variability exhibited by modern populations. This motivates us to examine the facial geometry of a large number of modern humans, of different races, different gender, and different age groups, to estimate the laryngeal position and its variability (Heim, Boë & Maeda, 2000). For this purpose, we use a biometric database describing the cranio-mandibular geometry of modern human populations developed at the Anthropology Laboratory of *Musée de l'Homme*, Paris. The database consists of measurements taken from midsagittal X-ray pictures of the heads of the contemporary modern human subjects mentioned just above, from dried crania of mummies from Egypt (again, male and female adults and children) and from the skulls of South American mummies, all of which allow for a comparison of populations from distinctly different areas. The method described was also applied to the skulls of two male adult Neandertals: *La Chapelle-aux-Saints* and *La Ferrassie I* (Heim, 1976) dated about 45,000 to 50,000 years. Both the cranium base and the mandible were preserved in these two fossils.

Some changes in the choice of landmarks were necessary, because the original landmarks, such as PNW (located on the soft tissues visible on the MRI data) are absent in our anthropological data (dried craniums of Egyptian mummies and of South American buried corpses). Consequently, we could only use landmarks on the bone structures. Moreover, the reconstructed skulls of Neandertals often lack the anterior nasal spine. We therefore attempted to determine OCH and PD from landmarks on the cranio-mandibular bone structures, namely, from the nasion-basion-gnathion triangle, as illustrated in Figure 1, with the help of correction factors to account for the difference in the landmarks. The correction factors are determined on the X-ray pictures of the head of contemporary human subjects (European male and female adults and children). Once values of OCH and PD are determined, we can calculate OHI, which is equal to OCH/PD by definition. Assuming the LHI-OHI relationship defined by the regression line of H&T is applicable to our data, LHI can be estimated directly from the determined OHI values.

The results are presented in Table I.

TABLE I. Range and mean values of Larynx Height Index estimated using the method proposed by Heim, derived from Honda & Tiede (with the number of subjects).

	Men	Women	Children
Modern (range)	0.54 – 1.00	0.60 – 0.85	0.60 – 0.74
Egyptian mummies	0.76 (10)	0.70 (5)	0.65(6)
South Amerindian mummies	0.80 (24)	0.73 (16)	0.68 (4)
<i>La Ferrassie 1</i>	0.71		
<i>La Chapelle-aux-Saints</i>	0.60		

Although we admit that our adaptation to the original H&T method could be improved, the estimated LHI values already exhibit a coherent trend. For the three distinctly different populations (modern, Egyptian, and South American) the average LHI value varies as largest for male adults (corresponding to the lowest laryngeal position), intermediate for female adults, and then smallest for children (the highest laryngeal position). These differences, linked to age and sex, have been extensively documented by Goldstein (1980). The estimated LHI values of Egyptian and South American groups exhibit fairly large dispersions, which appear to make the inter-group differences irrelevant. As far as the two male Neandertals are concerned, their LHI value is within the limits of the variation observed in modern populations. In detail, however, their LHI value is lower than the average values of individual male adult groups. In fact, the LHI value of the Neandertal man from *La Ferrassie 1* corresponds to the average value of female adults. The value for the man from *La Chapelle-aux-Saints* corresponds to the lower limit for the modern female adults and close to the average LHI of children. The particularly small LHI value of the man from *La Chapelle-aux-Saints* is due to his exceptionally large PD value.

LHI is therefore an important articulatory parameter, since it reflects growth differences from birth to adulthood and gender differences, as well as differences between homo sapiens and Neandertal. However, we shall demonstrate that this parameter has a minor influence on the realization of maximal vowel contrasts such as [i a u]; indeed, articulatory gestures of the tongue and lips allow compensation for these relative differences in vocal tract dimensions.

3. A growth model to simulate acoustical consequences of Larynx Height Index variations

We carried out articulatory and acoustic vocal tract simulations, with the following objectives: (i) prediction of the acoustic space corresponding to different values of LHI; (ii) inference of the maximal acoustic space potentially used by Neandertal men. The use of a vocal-tract model designed to study human growth seemed appropriate, since the difference in LHI between the newborn and the adult male is quite large, and the model therefore should account for a good deal of variability.

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 pediatry journals. 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. Goldstein (1980) 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, is provided for ages ranging from a few months to 20 years. All of these data can be closely fitted by (double) sigmoidal curves, which characterize the general patterns of skeletal and muscular growth. Here we summarize and draw attention to the points that are essential in understanding the phenomenon of vocal tract growth.

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 emphasized. For the vocal tract, this phenomenon is further accentuated by lowering of the larynx (inferred, in X-rays, from the position of the hyoid bone). The growth of the pharynx is therefore approximately twice as large as that of the front cavity. If we again turn to the data gathered by Goldstein, it is possible to recover an estimation of LHI and its evolution during growth (Figure 2) using the following landmarks and dimensions:

- the anterior Bjork's articulare (intersection of the posterior outline of the neck of the head of the mandibular condyle with the outline of the inferior surface of the occipital bone), the anterior nasal spine and the glottis.
- the horizontal distance from the anterior nasal spine to the articulare, and the vertical distance between the glottis and the reference line traced from the anterior nasal spine to the posterior nasal spine.

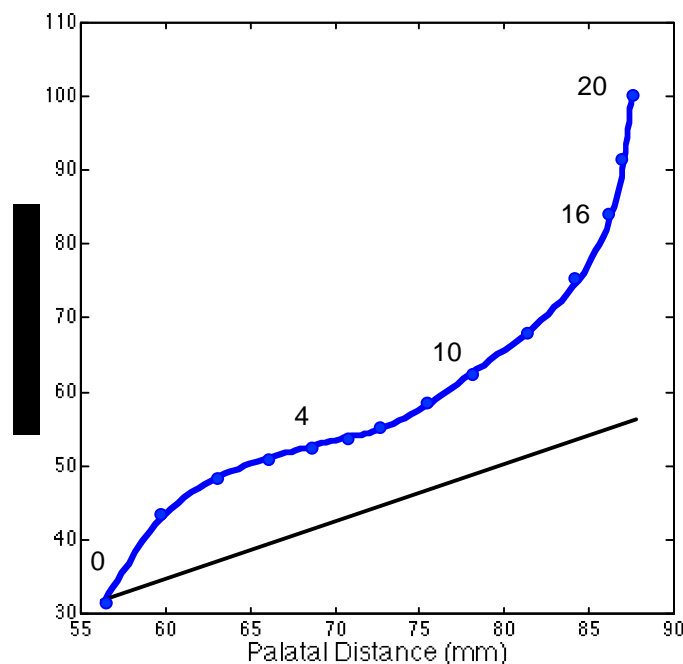


Figure 2. Variations of Larynx Height as function of Palatal distance. Points correspond to ages of 0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, and 20 years (from Goldstein, 1980). The straight line would correspond to a linear growth of the vocal tract.

To represent the vocal tract growth, the articulatory model of Goldstein describes the evolution of the horizontal and vertical dimensions from a newborn to a female or male adult. Here we have used the *Variable Linear Articulatory Model* (VLAM), developed by Maeda (cf. Boë, Maeda, 1998) on the basis of Maeda (1989). As proposed by Goldstein, 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, the other for the pharynx, with interpolation of the zone in-between.

The vocal-tract model does not allow the recovery of exactly the same landmarks as those used by Honda & Tiede, resulting in slightly higher values for LHI of the model, compared to LHI measured by Honda & Tiede. For the palatal distance PD and larynx height, we retained the incisors, the pharyngeal wall and the glottis position (Figure 3). For Neandertals, the value of PD has been estimated as the distance between the dental arch and the foramen magnum taken from a cast of the skull of the man from *La Chapelle-aux-Saints* rebuilt by Heim, hypothesizing that the larynx of the latter was in the same position as the one of the *Homo Sapiens Sapiens* (Table II).

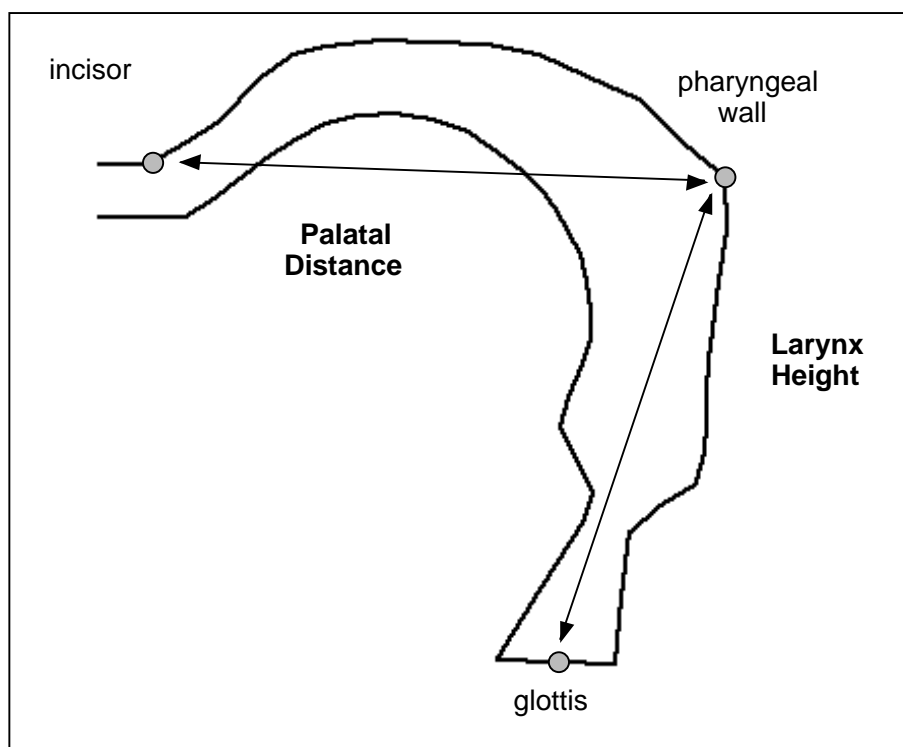


Figure 3. Reference points selected on the midsagittal curve generated by the model.

TABLE II. The value of the vertical larynx height (LH) and horizontal palatal distance (PD) dimensions calculated using the articulatory model (VLAM).

	LH	PD	LHI (model)
Neandertal (adult male)	8.80	10.0	0.88
21-year-old male	8.70	8.70	1.00
Adult female or 16-year-old male	7.40	7.80	0.95
10-year-old child	5.75	6.57	0.88
4-year-old child	4.51	5.70	0.79
0-year-old newborn	2.63	4.34	0.60

Furthermore, we ensured that length variation in the midsagittal dimension corresponds to data. We retained Goldstein's (1980) data, gathered from her model, as well as recent MRI data of Fitch & Giedd (1999) (and also Yang & Kasuya, 1994, for a boy of 11 years old, Story et al., 1996 for adults). It is worth noting that Goldstein's model, as well as VLAM, predict vocal tract lengths, during growth, that lie within the (maximum and minimum) limits presented by Fitch & Giedd (Figure 4). This assessment is of great importance, to the extent that vocal tract length determines the absolute position (along the frequency axis) in the maximal vowel space, and hence of the vowel formants.

Not surprisingly, the model generates satisfactory vocal tract lengths for adult men and women since it was established based on radiographic data for an adult woman, but it is important to verify the values for vocal tracts corresponding to children of 11 years of age and younger.

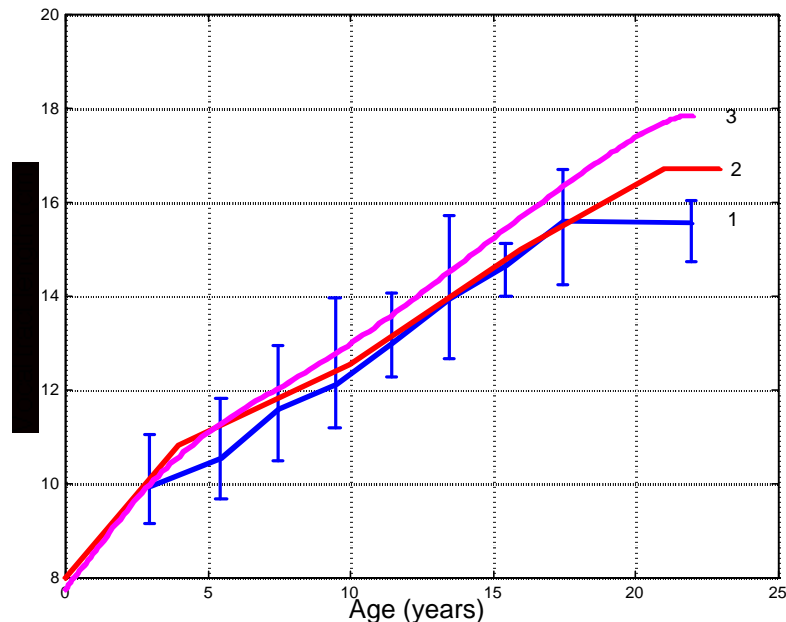


Figure 4. Variations in vocal tract length as a function of age: (1) minimum, mean and maximum values (Fitch & Giedd, 1999); (2) data proposed by Goldstein (1980); (3) data delivered by the articulatory model (VLAM).

4. Maximal Vowel Space and point vowels [i a u]

4.1. *The Maximal Vowel Space as a tool of prediction*

The question now is: what is the range of formant frequencies that can be produced by all possible shapes that correspond to a given vocal tract (a given LHI)? We define Maximal Vowel Space (MVS) as the n-dimensional space within which are situated the n first formants of all possible vocalic sounds that can be realized by a given vocal tract. For several years, the 3D space F_1 - F_2 - F_3 has been systematically measured (Peterson & Barney 1952) and modeled (Fant 1960). Obviously, the use of a realistic articulatory model allows a better evaluation of the limits of that space (Boë *et al.* 1989).

The value of each parameter of the articulatory model is constrained within a realistic range, chosen to correspond to three standard deviations (σ) either side of its mean value. If the entire input space of command parameters is explored while satisfying the conditions necessary for vowel production, one can simulate the maximal F_1 - F_2 - F_3 acoustic space at the output. All possible oral vowels are thus situated within the limits of this region. By systematically covarying all parameters in the model, compensatory effects can be taken into account.

This method allows for a precise description of all the possibilities for maximal distinctiveness and permits an optimal choice of prototypical realizations. Such an approach can be shown to be more reliable than one that consists of 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 L&C).

4.2. *Simulations of the MVS from human birth to adulthood and the Neandertal male vowel space*

In our simulations we immediately encountered the problem of the choice of the number of simulations and the values of the set of parameters to obtain a potential Maximum Vowel Space. We adopt, for each articulatory parameter, a random uniform distribution in the interval -3σ , $+3\sigma$ (except for larynx height: -1σ , $+1\sigma$ which is more realistic). Then we imposed minimum thresholds 0.3 cm^2 (*e.g.*, Fant, 1960; Catford, 1977) for constriction area which is the standard value discussed above; for the lip threshold this was lowered to 0.1 cm^2 , a value observed during speech production, especially for closed vowels such as [u] and [y] (Abry & Boë, 1986). Moreover, we also imposed a maximum possible cross-sectional area (*e.g.* 8 cm^2 for adult men) to avoid a too large lip opening. The minimum cross-sectional area requirement, therefore, constitutes the necessary condition for a specified vocal-tract configuration to be considered as that of a vowel. It may be noted that this minimum value is valid regardless of the global tract size.

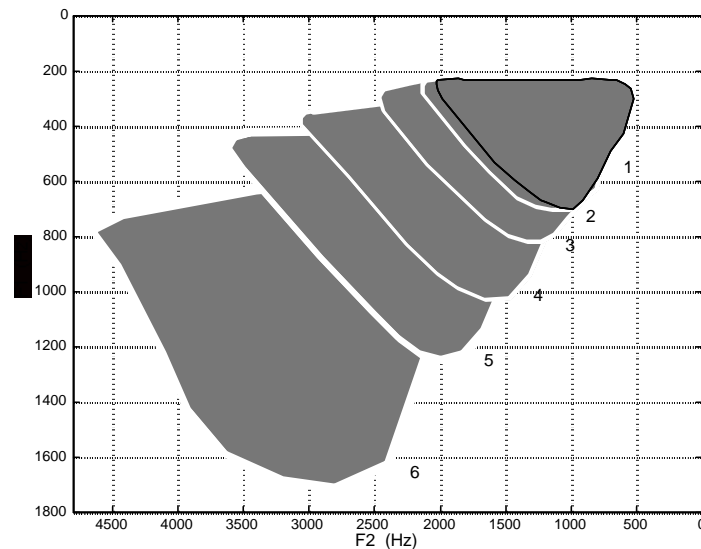
With a set of values specified for the seven parameters, vectors for the corresponding articulation are calculated. These vectors are projected onto the coordinates system, modified by the two scale factors, resulting in a vocal tract profile at a particular growth stage. The corresponding area function is then estimated by assuming that the cross-sectional area, at a given point along the tract, is expressed as a power function of the midsagittal dimension at that point (Heinz & Stevens, 1964; Perrier *et al.* 1992). As described above, if a cross-

sectional area of the area function becomes less than the minimum, or exceeds the maximum area value, the procedure stops at this point. If not, the frequency of the first three formants is calculated by solving acoustic equations governing the propagation of waves inside the tract (Badin & Fant, 1984). With the minimum and maximum cross-sectional area tests we generated 10,000 configurations for each vowel space. Average vocal tract lengths, standard deviations and variation coefficients, as a function of age are presented in Table III. According to data from Goldstein (1980), the vocal tract length of an adult female can be considered to correspond approximately to that of a 16-year-old male.

TABLE III. Average vocal tract lengths, standard deviations and variation coefficients, as a function of age.

Age (years)	0	4	10	Adult woman (16-year-old man)	Adult male	Neandertal (adult male)
Length (cm)	7.72	10.83	12.87	15.68	17.84	19.38
σ	0.33	0.44	0.53	0.65	0.76	0.86
σ /Length (%)	4.27	4.06	4.11	4.14	4.26	4.44

The cloud of all data points on the F_1 - F_2 plane indicates the potentially maximum vowel space for a given morphology, specified by the two scale factors of the front and back cavity lengths. The results are illustrated in Figure 5, where the potential Maximum Vowel Spaces are shown for a Neandertal male, an adult male, an adult female a child and a newborn.



Figures 5. Model-generated potential Maximum Vowel Space for (1) a Neandertal, (2) an adult male (21 years old), (3) an adult female (or a 16-year-old male), (4) a 10 year-old-child, (5) a 4-year-old child, and (6) a new-born infant (about one month old), shown on the F_2 - F_1 plane.

4.3. *The location of [i a u] inside the Maximal Vowel Space:
The flaw of Lieberman & Crelin*

Examining the shape of this maximal space, it is evident that constraints exist among the three first formants. If one considers, following Lindblom & Liljencrants (1974) and Schwartz, Boë, Vallée & Abry (1999b), that the vowels [i a u] are situated within the space so as to maximize the distances between vowels, it is possible to characterize the three basic vowels in the following way (Boë, Abry, Beautemps, Schwartz, & Laboissière, 2000). The vowel [i] is characterized by a maximal F_3 , the vowel [a] corresponds to a maximal F_1 , and the vowel [u] is produced with a minimal F_2 . Due to the shape of the maximal vowel space in the 3D F_1 - F_2 - F_3 space the following values are involved: the vowel [i] results in a minimum F_1 , the vowel [a] results in close F_2 and F_3 values, and the vowel [u] results in a minimal F_1 .

The following articulatory relations are well known:

[i] contains a large pharyngeal cavity with a narrow oral constriction;

[a] contains a narrow pharyngeal constriction and a large oral cavity;

[u] contains two large cavities of approximately equal size linked by a narrow constriction, and a second constriction at the lips (two Helmholtz resonators).

By using these acoustic criteria for the three extreme vowels, and by selecting among the 10,000 vowel samples, we obtain the values for the vocal tract length and area functions for F_1 and F_2 . If we compare the lengths, area functions and corresponding formants proposed by L&C (figure 28 of the paper published in 1971), with these data (Figure 6 and Table IV) it can be seen that:

- The vocal tract length of Neandertal estimated by L&C, corresponds to that of a modern woman. Since the palatal distance of Neandertal from *La Chapelle-aux-Saints* (estimated from the skull), is by about 2 cm greater than that of a modern man, this would mean that Neandertal had a larynx located 4 cm higher than the larynx of a modern man. Undoubtedly, this very high larynx position is unlikely;
- Constriction lengths are too short and constriction values are insufficiently small, so that L&C could only obtain lowered vowels corresponding to ['e'] and ['o'], rather than the extreme vowels [i] and [u];
- Finally the formants of the vowel [a] are not coherent with the Maximal Vowel Space of an adult female speaker (F_1 and F_2 are too high).

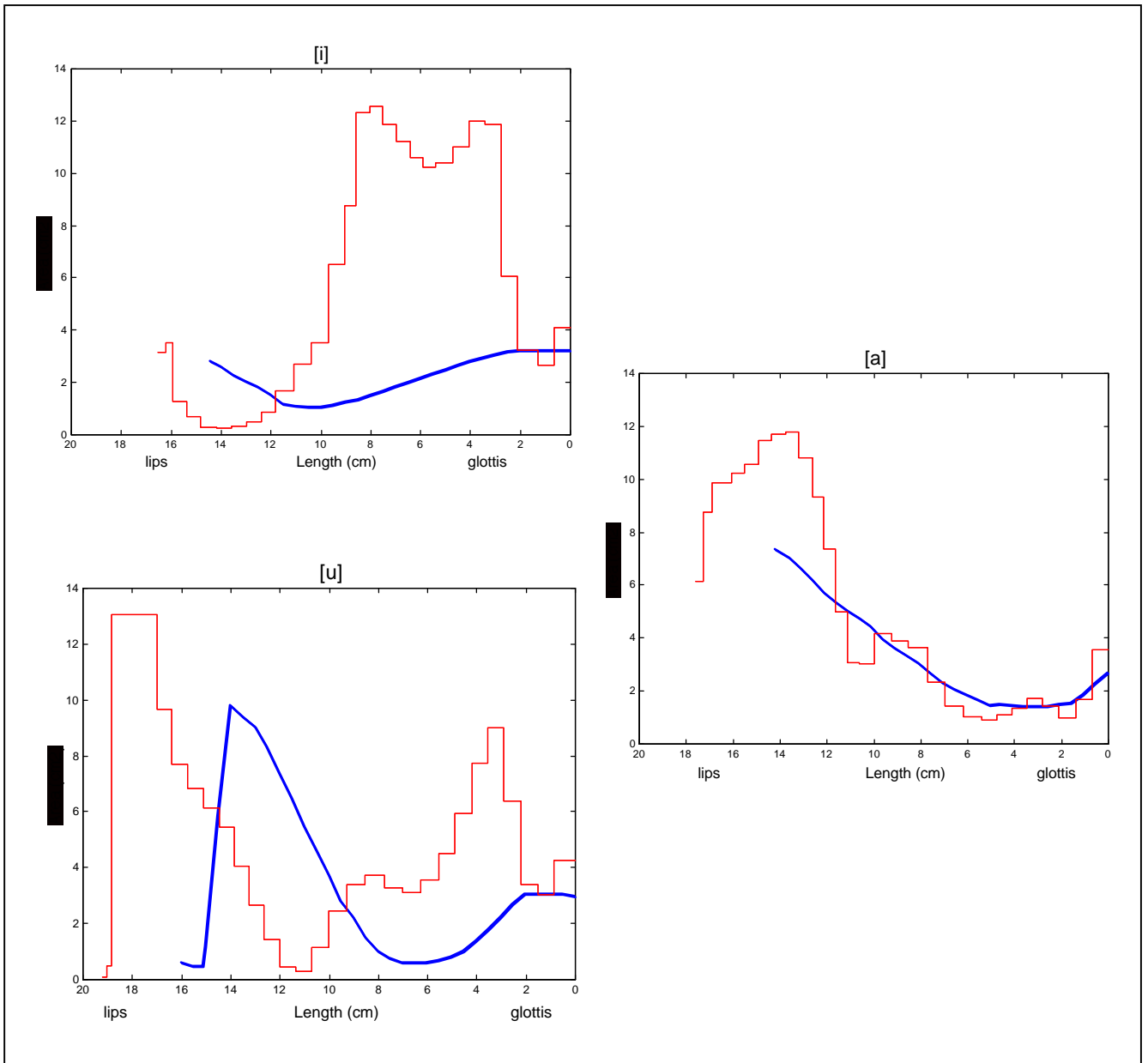


Figure 6. Comparison of area functions for [i a u]. Solid lines correspond to area functions proposed by Lieberman & Crelin (1971) for Neandertal, and staircase to typical area functions obtained from the Maximal Vowel Space for a hypothetical Neandertal vocal tract (adult male).

TABLE IV. Lengths of the vocal tract and the values of F_1 - F_2 for [i a u] proposed by L&C for the Neandertal male and for the Neandertal male adult and female adults estimated in this study.

	L (cm)	F_1 (Hz)	F_2 (Hz)
Neandertal male (Lieberman & Crelin)			
[i]	14.4	524	2038
[a]	14.2	846	1650
[u]	16.0	462	807
Neandertal male			
[i]	17.2	270	1995
[a]	18.	675	1095
[u]	19.5	260	600
Human female			
[i]	15.0	280	2200
[a]	14.8	780	1280
[u]	16.0	280	655

5 Normalization of different Maximal Vowel Spaces

"Uniform normalization by vocal tract length scaling" is a procedure employed systematically in the literature when F-patterns of a vowel from vocal tracts of different sizes are compared (*e.g.*, Lieberman & Crelin, 1971; Goldstein, 1980). Formant frequencies are scaled up or down in inverse proportion to the ratio of the total length of a given tract to that of a reference vocal tract. The difference in the corresponding formant frequencies after the normalization is considered to be due to the shape difference. Ultimately, the objective of speech communication is of course to "speak to be heard in order to be understood" (Jakobson, 1976), and hence vowel spaces should be described in terms of perceptual contrasts and auditory distances. It is well known that the perceptual frequency scale is more or less semi-logarithmic, i.e. linear below about 1 kHz and logarithmic above. A pure logarithmic scale would provide a transformation equivalent to a uniform length scaling, since a vocal tract derived from a reference shape by linear scaling would provide exactly the same auditory distances. Perceptual semi-logarithmic scales such as Bark (Schroeder, Atal & Hall, 1979) or erbs (Moore & Glasberg, 1983) expand the low-frequency region somewhat. In this region (typically, the F_1 region), smaller vocal tracts produce larger formant values, and hence increased perceptual contrasts. Indeed, we see for vowel spaces expressed in Bark (Figure 7) that only newborn infants display a vowel space that is slightly larger than that of adults in terms of F_1 distances. It is obvious that the potential Maximum Vowel Space does not vary much as a function of the growth stage and of gender and is therefore not sensitive to the relatively large differences in LHI. The relative closeness across different growth stages of the F_1 lower limit, corresponding the upper limit of the vowel space, regardless of growth stage should not be too surprising. The lower limit corresponds to F_1 for the vowels [i] and [u]. Since F_1 of these vowels is associated with a Helmholtz resonance, the value can be lowered to the limit by narrowing the neck of the resonator by appropriate articulatory maneuvers. The closeness of the F_1 upper limit, corresponding to the bottom of the vowel spaces occupied by open vowels such as [a], deserves explanation. Both F_1 and F_2 of an open vowel are

associated with the quarter-wavelength resonance mode where the frequency is determined by cavity length. It might be expected then that F_1 and F_2 frequencies are directly influenced by the morphological characteristics of the tract. The closeness of the F_2-F_1 dispersion regardless of growth stages suggests that the morphological differences are compensated by an articulatory maneuver. We speculate, as Goldstein (1980) has already suggested, that a newborn having an extremely long front cavity relative to the back cavity can produce the extreme open vowel [a] by raising the back of the tongue, resulting in a lengthening of the pharyngeal cavity concomitant with a shortening of the oral cavity. Such a compensatory maneuver must be possible, at least in theory, by activating the styloglossus muscles. With the model, assigning a higher value to the tongue-body shape parameter could simulate this maneuver. In the case of a child, the length asymmetry is much less severe than that of the newborn.

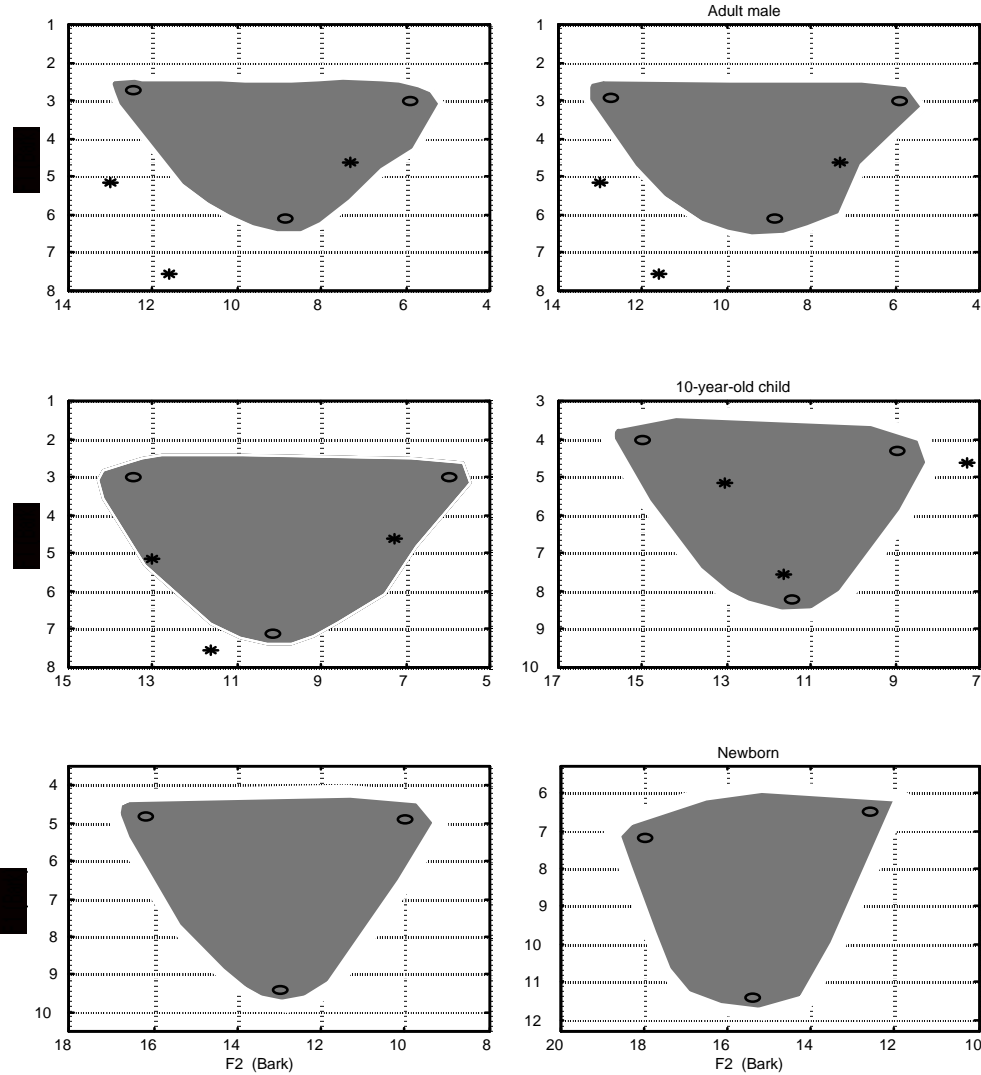


Figure 7. Maximal Vowel Spaces normalized by a perceptual scale in Bark and the values of F_1-F_2 for [i a u] proposed by Lieberman & Crelin for the Neandertal male (*), and for the Neandertal male, male adults and female adults estimated in this study (o).

Considering that the estimated LHI of the Neandertal man from *La Chapelle-aux-Saints* is close to that of a child, it is not so unreasonable to state that the Neandertal could have had a normalized vowel space comparable in size to that of human adult speakers. As in the case of the child, a relatively long oral cavity relative to the pharyngeal cavity requires some degree of articulatory compensation. On the basis of topologically similar anatomical arrangements of the extrinsic tongue muscles of mammals and of primates including humans, Neandertals probably had a muscular arrangement similar to that of modern humans. If this is the case, the Neandertal male could have had a comparable vowel space. As far as the Neandertal man from *La Ferrassie 1* is concerned, his LHI is close to that of an adult female speaker. Consequently, he did not, at least morphologically, have any trouble producing any human vowel.

6. Why equivalent Maximal Vowel Space with different Larynx Height Index?

Normalization results tend to show that, no matter what the ratio between oral and pharyngeal cavities, vowel spaces are approximately similar. What are the acoustic effects of varying the ratio of the back to the front cavity length? The effects are clearest for vowels such as [a], because both F_1 and F_2 frequencies are functions of the cavity length. F_1 and F_2 are closest when the two cavities have the same length, *i.e.* the ratio is equal to one (see *e.g.* Maeda, Carré, 1995). This configuration occurs between the average configurations for adult male and adult female speakers. An adult male speaker tends to have a longer back than front cavity, *i.e.* the ratio is greater than one. After normalizing the female and male vocal tract lengths, we should observe that F_1 and F_2 of the vowel [a] produced by the male and female speakers are of about the same frequency. When the ratio becomes smaller than that of the adult female, corresponding to the tract configuration of a child or of a Neandertal male, F_1 associated with the front cavity goes down in frequency and F_2 , associated with the back cavity, goes up, losing somewhat its phonetic value as the vowel [a]. Indeed, for the extreme configuration of the baby, the parameter values for the adult [a] can even result in an [æ]-like vowel in the articulatory model.

In the extreme, the ratio can fall below 0.5, *i.e.* such that the front cavity length is more than twice the back cavity length, for the tract morphology for a newborn infant. In this case, F_1 and F_2 are so far apart that the vowel produced will no longer be identified as the vowel [a] but rather as something like [E]. Indeed this simplified analysis appears to suggest that the newborn infant cannot produce the important corner vowel [a]. In more general terms, the normalized acoustic space of newborns could be smaller than that of children and adults, *i.e.* the shape factor can influence the size of normalized vowel spaces.

We implicitly assumed that an underlying articulatory configuration for the vowel [a] remains invariant, regardless of tract morphology, and of course, of the total tract length. That is to say, the articulated vocal tract shape is deformed in accordance with a change in the tract morphology specified in terms of the length ratio. Contrary to the assumption, it is a common observation that the ways a vowel is uttered vary greatly depending on individual speakers. This articulatory variability could be explained, in part, by the fact that the individual speakers adapt the articulation of the same vowels to compensate for morphological differences of their vocal tract. For example, we mentioned already that formant frequencies essentially determined by the Helmholtz resonance (F_1 of [i], and F_1 and F_2 of [u]), can be arbitrarily manipulated by adjusting the constriction size. The consequence of the manipulation can actually be observed when we compare the vowel system of a language, plotted on the F_2 - F_1 plane, produced by female and male speakers (*e.g.*, Koopmans-Van Beinum, 1980: 54; Calliope 1989: 85; Lee, Potamianos & Narayanan, 1999). Non-normalized

F-patterns of vowels often indicate that data points for female and male [i] overlap each other along the F_1 axis, but not along the F_2 axis. For the vowel [u], as expected, the female and male dispersions in F_1 and F_2 overlap with each other. It is therefore reasonable to state that speakers articulate the same vowels differently, presumably to meet perceptual requirements (Nordström, 1977). This then implies that, when we compare the articulated vocal tract shapes of different speakers, we must take into account not only the effects of the neutral vocal tract morphology, but also the possible articulatory compensations. This is the primary reason for our calculations of the potential maximum vowel spaces using an articulatory model to compare the vowel production capability of speakers having different tract morphologies.

Recent simulations have confirmed the compensation of the tongue body parameter in the model for the newborn realization of [a] (Ménard & Boë, 2000). For the two other vowels [i] and [u], the general tendency for a shorter vocal tract corresponding to a newborn infant and a child would consist in an advanced tongue position as well as closing of the lips (for [i]) and opening of the lips (for [u]). Note that these results were obtained by an inversion procedure and by choosing the parameter values which involved minimal articulatory perturbation compared to the adult speakers. Our results can be predicted by formant-cavity affiliations for each vowel.

7. Conclusions

Modeling the growth of the vocal tract enables a better understanding of the phenomena governing anatomical differences between neonates, babies, adolescents, and male and female adults quantified by the Larynx Height Index. It allows discussion of the consequences of variation in vocal tract dimensions during evolution with the aim of establishing distinctive sounds for speech. It opens up new operational perspectives for normalization procedures.

For the time being, considering the data gathered, and our modeling results, it is reasonable to consider that the base of the skull, the hyoid bone, the position of the larynx, and the dimension of the pharynx of Neandertals were the same as those of modern humans; due to a larger Palatal Distance their Larynx Height Index was lower than that of modern humans.

However these important anatomical considerations have limited consequences for the ability for Neandertals to contrast [i a u]. Our simulations show that the Maximal Vowel Space of a given vocal tract does not depend on the Larynx Height Index: gestures of the tongue body (and lips and jaw) allow compensation for differences in the ratio between the dimensions of the oral cavity and pharynx.

These results confirm the conclusions of Goldstein, 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 realizing vowel contrasts!

The descent of larynx is generally attributed to the upright posture of mankind (Wind 1983). But a low larynx (and large pharynx) can not be considered to be the "*anatomical prerequisites for producing the full range of human speech*", and there is no reason to believe that lowering of the larynx and an increase in pharynx size are necessary evolutionary pre-adaptations for speech. 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. According to present evidence, monkeys can not talk 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 for maximally contrasting vowels.

If Neandertals could not talk, it is unlikely to have been for the articulatory acoustic reasons advocated by L&C. Neandertals were no more vocally handicapped than children at birth are (Figure 8).

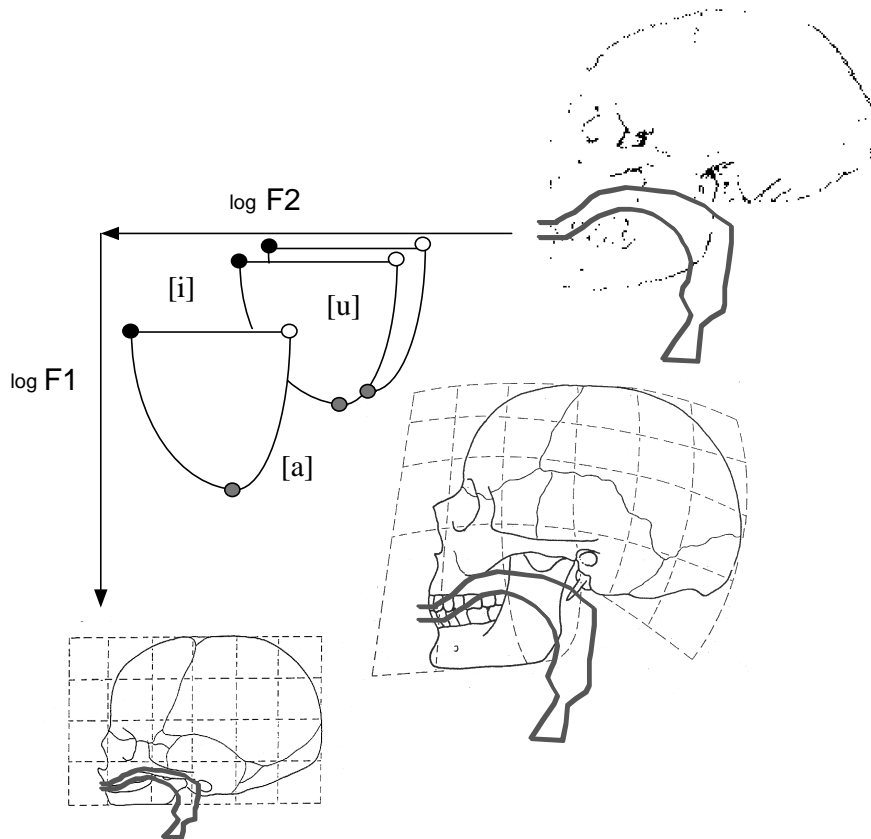


Figure 8. Maximal Vowel Spaces (schematized and normalized) for newborn and adult Homo Sapiens Sapiens, and a hypothetical Maximal Vowel Space for a Neandertal male.

We are grateful to Christian Abry, Denis Autesserre, Pierre Badin, Denis Beautemps, Barabara Davis, Didier Demolin, Rafael Laboissiere, Peter MacNeilage, Lucie Ménard, Gordon Ramsay, Solange Rossato, Philip Rubin, Jean-Luc Schwartz, Marija Tabain, Marilyn Vihman, Phil Hoole, the two anonymous reviewers for their helpful criticisms, data, suggestions, translation, corrections and encouragement.

References

- Abry, C. & Boë, L.J. (1986) "Laws" for lips, *Speech Communication*, **5**, 97-104.
- Arensburg, B., Bar-Yosef, O., Chech, M., Goldberg, P., Laville, H., Meignen, L., Rak, Y., Tchernov., E., Tillier & A.-M., Vandermeersch, B. (1985) Une sépulture néandertalienne dans la grotte de Kébara (Israël), *Compte Rendu de l'Académie des Sciences de Paris* **300**(II, 6), 227-230.
- Arensburg, B., Tillier, A.-M., Vandermeersch, B., Duday, H., Schepartz, L.A. & Rak, Y. (1989) A middle paleolithic human hyoid bone, *Nature*, **338**, 758-760.
- Badin, P., Fant, G. (1984) Notes on vocal tract computations, *STL-QPSR*, **2-3**, 53-108.
- Boë, L.-J. (1999) Modeling the growth of the vocal tract vowel spaces of newly-born infants and adults: Consequences for ontogenesis and phylogenesis. In *Proceedings of the 14th International Congress of Phonetic Sciences*, **3**, pp. 2501-2504.
- Boë, L.-J., Abry, C., Beautemps, D., Schwartz, J.-L. & Laboissière, R. (2000) Les sosies vocaliques. Inversion et focalisation, *XXIIIe Journées d'Étude sur la Parole, JEP'2000*, 257-260.
- 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.
- Boë, L.-J., Maeda, S., Abry C. & Heim, J.-L. (1998) /i a u/ ? À la portée d'un conduit vocal de Neandertal, In *Actes des XXII^e Journées d'Études sur la Parole*, Martigny, Suisse, 245-248.
- Boë, L.-J., Perrier, P., Guerin, B. & Schwartz, J.-L. (1989) Maximal vowel space, In *Proceedings of Eurospeech*, **2**, 281-284.
- Boule, M. (1913) L'Homme fossile de La Chapelle-aux-Saints, *Annales de Paléontologie humaine*, Paris: Masson.
- Boule, M. (1921) *Les hommes fossiles. Éléments de paléontologie humaine*. Paris: Masson.
- Calliope (1989) *La parole et son traitement automatique*. Paris: Masson.
- Catford, J.C. (1977) *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.
- Falk, D. (1975) Comparative anatomy of the larynx in man and the chimpanzee: implications for language in Neanderthal, American, *Journal of Physical Anthropology* **43**, 1, 123-32.
- Fant, G. (1960) *Acoustic theory of speech production*. The Hague: Mouton.
- Fant, G. (1966) A note on vocal tract size and non-uniform F-pattern scalings, *STL-OPSR* **4**, 20-30.
- Fitch, W.T. & Giedd, J. (1999) Morphology and development of the human vocal tract: A study using magnetic resonance imaging, *Journal of the Acoustical Society of America*, **106**, 1511-1522.
- Goldstein, U.G. (1980) *An articulatory model for the vocal tract of the growing children*. Thesis of Doctor of Science, MIT, Cambridge, Massachusetts.
<http://theses.mit.edu:80/Dienst/UI/2.0/Describe/0018.mit.theses%2f1980-17?abstract=>
- Heim, J.-L. (1976) Les Hommes fossiles de La Ferrassie. Tome I. Le gisement, les squelettes adultes : crâne et squelette du tronc, *Archives de l'Institut de Paléontologie Humaine*, 35. Paris: Masson.
- Heim, J.-L. (1986) La reconstitution du crâne de la Chapelle-aux-Saints, Film 16mm. Paris: Éditions SFRS.
- Heim, J.-L. (1989) Une nouvelle reconstitution du crâne néandertalien de la Chapelle-aux-Saints, *Compte Rendu de l'Académie des Sciences de Paris*, **308**(II, 6), 1187-1192.
- Heim J.-L. (1990) La nouvelle reconstitution du crâne néandertalien de la Chapelle-aux-Saints. Méthode et résultats, *Bulletin et Mémoires de la Société d'Anthropologie de Paris*, **6**(1-2), 94-117.
- Heim, J.-L., Boë, L.-J. & Maeda, S. (2000) Essai de la détermination de la position du larynx à partir de repères craniométriques: Application à la paléontologie humaine. In *L'identité humaine en question. Nouvelles problématiques et nouvelles technologies en Paléontologie humaine et en Paléonthropologie biologique* (Ph. Andrieux, D. Hadjouis, A. Dambricourt-Malassé, editors), pp. 187-204, Paris: Artcom.
- Heinz, J.M. & Stevens, K.N. (1964) On the derivation of area functions and acoustic spectra from cineradiographic films of speech, *Journal of the Acoustical Society of America*, **23**, 1037-1038.
- Hillenbrand, J., Getty, L.A., Clark, M.J. & Wheeler, K. (1995) Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America*, **97**, 3099-3111.
- Honda, K. & Tiede, M.K. (1998) An MRI study on the relationship between oral cavity shape and larynx position, In *Proceedings of the 5th International Conference on Spoken Language Processing*, **2**, 437-440.
- Houghton, P. (1993) Neandertal supralaryngeal vocal tract, *American Journal of Physical Anthropology*, **90**, 139-146.
- Jakobson, R. (1976) *Six leçons sur le sons et le sens*. Paris: Minuit.

- Kasuya, H., Suzuhi, H. & Kido, K. (1968) Changes in pitch and first three formants of five Japanese vowels with age and sex speakers, *Journal of the Acoustical Society of Japan* [in Japanese] **24**, 355-364.
- Koopmans-Van Beinum, F.J. (1980) *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers. B.V.
- Kuhl, P. & Meltzoff, A.N. (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change, *Journal of the Acoustical Society of America*, **100**, 2425-2438.
- Laitman, J.T. (1983) The evolution of hominid upper respiratory system and implication for the origin of speech, *Glossogenetics*, 63-90. Harwood Academic publishers.
- Laitman, J.T., Heimbuch, R.C. & Crelin, E.S. (1979) The basicranium of fossil hominids as an indicator of their upper respiratory systems, *American Journal of Physical Anthropology* **51**, 15-34.
- Lee, S., Potamianos, A. & Narayanan, S. (1999) Acoustics of children's speech: Developmental changes of temporal and spectral parameters, *Journal of the Acoustical Society of America*, **105**, 1455-1468.
- Lieberman, D.E. & McCarthy, R.C. (1999) The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution*, **36**(5), 487-517.
- Lieberman, Ph. (1972) *The speech of primates*. The Hague: Mouton.
- Lieberman, Ph. (1973) On the evolution of language: A unified view, *Cognition*, **2**, 59-94.
- Lieberman, Ph. (1984) *The biology and evolution of language*. Cambridge, Mass: MIT Press.
- Lieberman, Ph. (1991) *Uniquely human. The evolution of speech, thought, and selfless behaviour*. Cambridge, Mass: Harvard University Press.
- Lieberman, Ph. (1994) Functional tongues and Neanderthal vocal tract reconstruction: a reply to Dr. Houghton (1993), *American Journal of Physical Anthropology*, **95**, 443-450; discussion 450-452.
- Lieberman, Ph. & Crelin, E.S. (1971) On the speech of the Neanderthal man, *Linguistic Inquiry* **2**, 203-222.
- Liljencrants, J. & Lindblom B. (1972) Numerical simulations of vowel quality systems: The role of perceptual contrasts, *Language*, **48**, 839-862.
- Maddieson, I. (1986) *Pattern of sounds*. 2nd edition, Cambridge, Mass.: Cambridge University Press (1st edition 1984).
- Maeda, S. (1989) Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In *Speech Production and Modelling* (W.J. Hardcastle & A. Marchal, editors) pp. 131-149: Academic Publishers, Kluwer.
- Maeda, S. & Carré, R. (1995) Modèles de production. In *École Thématique Fondements et perspectives en traitement automatique de la parole*, GDR-PRC Communication Homme-Machine(H. MéLoni, editor), Université d'Avignon et des Pays du Vaucluse, pp. 51-72.
- McCarthy, R.C. & Lieberman, D.E. (1997) Reconstructing vocal tract dimensions from cranial base flexion : An ontogenetic comparison of cranial base angulation in humans and chimpanzees. *American Journal of Physical Anthropology Suppl.*, **12**, 163-164.
- Ménard, L. & Boë, L.-J. (2000) Exploring vowel production strategies from infant to adult by means of articulatory inversion of formant data, In *Proceedings of the International Congress of Spoken Language Processing*, Beijing (China), 465-468.
- Mol, H. (1970) *Fundamentals of Phonetics II: Acoustical modes generating the formants of the vowel phonemes*. The Hague: Mouton.
- Moore, B.C.J. & Glasberg B.R. (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America* **74**, 750-753.
- Nordström, P.-E. (1977) Female and infant vocal tracts simulated from male area functions, *Journal of Phonetics*, **5**, 81-92.
- Perrier, P., Boë, L.-J. & Sock, R. (1992) Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modeling the transition with two sets of coefficients, *Journal of Speech and Hearing Research*, **35**, 53-67.
- Peterson G.E. & Barney, H.L., (1952) Control methods used in a study of vowels, *Journal of the Acoustical Society of America*, **24**, 175-184.
- Schepartz, L.A. (1993) Language and modern human origins, *Yearb. Physical Anthropology*, **36**, 91-126.
- Schwartz J.L., Boë L.J., Vallée N. & Abry C. (1997a) Major trends in vowel system inventories, *Journal of Phonetics*, **25**, 233-253.
- Schwartz J.L., Boë L.J., Vallée N. & Abry C. (1997b) The Dispersion-Focalization Theory of vowel systems, *Journal of Phonetics*, **25**, 255-286.

- Schroeder, M.R. Atal, B.S. & Hall, J.L. (1979) Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In *Frontiers in speech communication research* (B. Lindblom & S. Öhman, editors), pp. 217-229, London: Academic Press.
- Smith, B.L. & Kay Kenney, M. (1998) An assessment of several acoustic parameters in children production development: longitudinal data, *Journal of Phonetics*, **26**, 95-108.
- Story, B.H., Titze, I.R. & Hoffman, E.A. (1996) Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, **100**, 537-554.
- Trinkaus, E. & Shipman, P. (1993) *The Neandertals. Changing the image of mankind*. New York: Alfred A. Knopf.
- Vallée, N. (1994) *Systèmes vocaliques : de la typologie aux prédictions*, Doctoral thesis, Sciences du Langage, Université Stendhal: Grenoble, France.
- White, P. (1998) Formant frequency analysis of children's spoken and sung vowels using sweeping fundamental frequency production, *Speech, Music and Hearing, Quarterly Progress and Status Report, TMH-QPSR*, **1-2**, 43-52.
- Wind, J. (1983) Primate evolution and the emergence of speech. In *Glossogenesis. The Origin of evolution of language and speech* (E. de Grolier, A. Lock, C.R. Peters & J. Wind editors), 15-35, Harwood Academic Publishers
- Yang, C.-S. & Kasuya, H. (1994) Accurate measurement of vocal tract shapes from magnetic resonances images of child, female and male subjects, In *Proceedings of the International Congress of Spoken Language*, Yokohama (Japan), 623-626.