Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model

Lucie Ménard, Jean-Luc Schwartz, Louis-Jean Boë, Jérôme Aubin

Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model

Lucie Ménard[a] Jean-Luc Schwartz[b], Louis-Jean Boë[b], and Jérôme Aubin[a]

[a] *Université du Québec à Montréal, Département de linguistique et de didactique des langues, C. P. 8888, Succursale Centre-Ville, Montréal, Canada, H3C 3P8*

[b] *ICP, UMR CNRS No. 5009, INPG-Université Stendhal, INPG, 46 Avenue Félix Viallet, 38031 Grenoble Cedex 1, France*

Suggested running title: Articulatory-acoustic relationships during growth

Abstract

This paper reports on the articulatory-acoustic relationships involved during vocal tract growth. Data were taken from a database of ten French vowels uttered by 15 speakers ranging in age from 3 years old to adulthood. Despite the important acoustic variation encountered, one feature is displayed by all the speakers: the production of extreme focal vowels /i/, /u/, /a/, and /y/, realized with a strong concentration of spectral energy related to the proximity of two formant peaks. This feature represents an acoustic goal guiding the speaker's task. Our simulations using an articulatory model demonstrate that the realization of the focalization feature may require different articulatory gestures for young children compared to adults, consisting of adaptive articulatory strategies exploited to compensate for the small pharynx of the former. Perceptual tests show that achieving focalization results in a lower intelligibility for the children than for the adults. Due to the relatively shorter pharyngeal cavity of the child compared to the adult, focalization can not be achieved together with the perceptual objective related to rounded vowels /y/ in French. Results are discussed in light of the dispersion-focalization theory and the perception for action control theory (PACT).

1. Introduction

Non-uniform vocal tract growth has been an object of study for decades. It is well known that the speech mechanism undergoes important modifications during development (Vorperian *et al.*, 2005; Vorperian, 2000; Fitch and Giedd, 1999; Goldstein, 1980; Kent, 1976). At the physiological level, Vorperian *et al.* (2005) computed growth curves for the overall vocal tract length based on MRI (Magnetic Resonance Imaging) in subjects from birth to 6 years and 9 months, and in adults. Based on these curves, a newborn vocal tract is 7.1 cm long and a 4-year-old vocal tract is 10.5 cm long. The mean length of the adult vocal tract is about 16 cm for male and 14.3 cm for female. The ratio of the pharynx length to the oral cavity length increases during growth, corresponding to approximately 0.5 for infants and to approximately 1.1 for adult males (Goldstein, 1980). This phenomenon is referred to as “non-uniform vocal tract growth”. Furthermore, the sequential emergence of motor control gives rise to limited articulatory capabilities for children (Kent, 1976; MacNeilage, 1997; MacNeilage and Davis, 2000; Lacerda, 2003). Indeed, at the first stage of canonical babbling, the sole control of the jaw allows the production of bilabial, alveolar, or velar stops (/b d g/) mainly co-occurring with open and central vowels /a æ œ/. Cognitive and perceptual abilities are also affected (Green *et al.*, 2000; Kuhl, 1992; Nittrouer, 1992). All these transformations shape the vowel system produced by the speaker during vocal tract growth. An analysis of acoustic variability raises the question of invariance. Is the evolving speech production system guided by articulatory or acoustic patterns? A few studies address this issue, directly or indirectly. For instance, in the search for normalization factors, Fant (1975) and Nordström (1975) advanced scaling factors that could explain the acoustic variation observed for vowels produced by men, women, and children. Most of the time, even though a precise acoustic or geometrical modeling was attempted, cavity length alone could not explain all the encountered acoustic variation. Fant (1975) proposed different articulatory strategies used to realize the same vowel, due to different vocal tract configurations. These differences could be perceptually motivated.

With the use of articulatory modeling, the acoustic effects of non-uniform vocal tract length can be predicted and compared to observed variability. Thus, the general trends associated with non-uniform growth of the cavities (front cavity and pharynx) can be described and, most importantly, the effects of articulatory gestures examined for various growth stages. A preliminary study, based on articulatory modeling, of similar articulatory positions from birth to adulthood for the vowels /i y u a/ showed that the relative position of the vowels in the acoustic space was very different between the young child and the adult male, especially for /y/ (Ménard and Boë, 2000). The question arising is then the necessity of articulatory compensation strategies, to adapt for the shorter pharynx in infants. If we consider similar articulatory settings for the two speakers, are the perceptual goals reached? Or on the contrary, are compensation strategies required in order to reach this goal?

These questions are of major importance regarding the issue of production and perception relationships. According to Locke (1983), from the early stages of language acquisition, production-perception links are encoded by the speaker and specify the speech task, so that a given articulatory strategy is associated with specific acoustic patterns. During childhood, the gradual mastery of motor control as well as anatomical modifications constitute enormous transformations that greatly affect these relationships. As a result, achieving a perceptual template may require a recalibration of articulatory-acoustic links.

In this paper, we demonstrate that the production of French oral vowels, from childhood to adulthood, is guided by acoustic targets in the F1, F2 and F3 space which seem to require different articulatory strategies during the course of development. We use the simulations of an articulatory model (VLAM) that integrates non-uniform vocal tract growth to re-interpret the acoustic variability and stability encountered in data from natural productions and the organization of vowel space. The acoustic effects of the growth phenomenon, combined with variations of articulatory settings, provide a grid for analysis.

2. Theoretical background: Predicted effects of vocal tract growth on vowel production

The following section provides a brief description of the articulatory configurations involved in the production of French oral vowels by adult speakers. The description focuses on the extreme vowels /i/ (as in French "fit" – English "did"), /y/ (as in French "fut" – English "was"), /u/ (as in French "fou" – English "crazy"), and /a/ (as in French "fa" – English musical note "F"), since these vowels are produced with a small constriction area. Coupling effects are thus minimal and formant-cavity affiliations can be considered. Based on previous studies, we then discuss the effects of non-uniform growth of the cavities on the acoustic pattern related to these vowels. The formant-cavity affiliations provided below are based on Fant's (1960) 4-tube model, Stevens' (1989) quantal theory, and Boë's (1999) model of vocal tract vowel space growth. Note that these descriptions are simplified for the sake of clarity.

2.1. A qualitative description of the acoustic consequences of the gestures involved in the vowels /i/, /y/, /u/, /a/ for adults

The articulatory gestures underlying the /i/ versus /y/ contrast in French for adult speakers are well known (Abry and Boë, 1986). For /i/, besides spreading of the lips, the tongue is in a high and front position, creating a wide back cavity including the pharynx, and a narrow front cavity formed by the constriction of the tongue towards the front part of the palate[1]. The front and back cavities act as simple tube resonators and their resonant frequencies are very close to the formants. The configuration created by the whole vocal tract corresponds to a Helmholtz resonator and is affiliated to F_1. F_2 is the half wavelength resonance of the back cavity, and F_3 is the half wavelength resonance of the front cavity (Fant, 1960).

Compared to /i/, the basic gesture associated with the vowel /y/ is rounding/protrusion of the lips, the tongue still being in a high and front position. Such a movement of the lips lengthens the front cavity, resulting in a decrease of the affiliated formant[2] (F3 for /i/ and F2 or F3 for /y/). If the front cavity remains shorter than the back cavity, F_2 does not change and stays affiliated to the half wavelength resonance of the back cavity while F_3, the half wavelength resonance of the front cavity, decreases. However, if the front cavity becomes longer than the back cavity, its resonant frequency becomes lower than that of the back cavity. In such a case, both F_2 and F_3 decrease from /i/ to /y/, and in /y/, F_2 is affiliated to the front cavity and F_3, to the back cavity (Schwartz *et al*., 1993).

For /u/ (as in French “fou” – English “mad”), the first formant is affiliated to the Helmholtz resonator created by the front cavity and the labial tube, whereas the second formant is affiliated to the second Helmholtz resonator created by the back cavity and the constriction. F3 is related to the quarter wavelength resonance of the back cavity (Ménard and Boë, 2000) [3].

Finally, for an adult male, the first and third formants of the low vowel /a/ (as in French “fa” – English musical note “F”) are affiliated to the first and second resonances of the back cavity, whereas the second formant is affiliated to the front cavity. A schematic representation of these formant patterns is depicted in Figure 1. The dotted lines correspond to adult values whereas the solid lines represent the child values.

[Insert Figure 1 about here]

2.2. The influence of non-uniform vocal tract growth: a qualitative analysis

As previously mentioned, the adult’s vocal tract is not a uniform scaled up version of the infant’s vocal tract. At birth, the infant has a very short pharynx compared to the length of the oral cavity, whereas the pharynx for the adult male is comparatively much longer, and roughly of the same size as the oral cavity (Goldstein, 1980). For the adult female, the pharynx is still shorter than the oral cavity, but the difference is less acute than for the infant. These cavity length differences have important effects on the resulting values of the resonant frequencies, as depicted in Figure 1. Indeed, for a Helmholtz resonator, this frequency can be calculated by the following formula: $F = (c/2\pi)\sqrt{(A_{co}/L_{co}V_{ca})}$, where c is sound velocity (about 350 m/s), A_{co} is the constriction area, L_{co} is the constriction length, and V_{ca} is the cavity volume. As for single tube resonators, the n^{th} half-wavelength and n^{th} quarter-wavelength resonant frequencies of a tube correspond respectively to $F= nc/2L_{ca}$ and $F= (2n-1)c/4L_{ca}$, where L_{ca} is the cavity length (Stevens, 1989). As can be predicted by the formulae presented above, the shorter the cavity, the higher the formant. The effects of growth can be observed by comparing the solid line (child) and the dotted line (adult) in Figure 1. Considering the two vowels /i/ and /y/ described in the previous section, all the articulatory gestures remaining unchanged, it can be predicted that the difference between formant values affiliated to the back cavity from the child to the adult will be greater than the difference between formant values affiliated to the front cavity. As a result, for /i/, F2 increases more than F3 in the child productions, hence decreasing the difference between F2 and F3 and increasing the difference between F3 and F4[4].

In the case of /y/, most of the front-cavity lengthening due to lip rounding involves a decrease of the resonances affiliated to the front cavity, those values being much lower than those of the back cavity. As a result, F3 increases more than F2 in the child productions, hence the difference between F2 and F3 in /y/ should be greater for the child, compared to the adult male. Figure 1 schematizes the cases discussed above.

Figure 1 also shows the formant-cavity affiliations for /u/ and /a/. It appears that, besides overall increase of all formant frequencies, the distance patterns between F1, F2, and F3 for /u/ remain similar for the child-like vocal tract and the adult-like vocal tract. Thus, a small pharyngeal cavity compared to the front cavity, for the child, does not affect the formant ratios. This pattern can be related to the fact that F1 and F2 are affiliated to Helmholtz resonators, for which the resonant frequency is affected not only by cavity length but also by cavity volume, constriction length, and constriction area. Finally, for /a/, compared with the adult male, the increase of F2, affiliated to the back cavity, for the young child, is much greater than the increase of F1 and F3, affiliated to the front cavity. As a result, F1 and F2 are farther apart for the child than for the adult.

Altogether, these predictions suggest that among the four vowels /i/, /y/, /u/, and /a/, the vowels /i/, /y/, and /a/ are more likely to be affected by non-uniform modifications of the vocal tract cavities. Indeed, compared to the adult male vocal tract, the F3-F4 distance for /i/, the F2-F3 distance for /y/ and the F1-F2 distance for /a/ are increased for a child-like vocal tract, /. Since these patterns, referred to as focalization (Abry *et al.,* 1989), have been found to be a criteria of local stability in vowel systems of the world's languages (cf the Dispersion-focalization theory of vowel systems, Schwartz *et al.*, 1997, 2005), it is worth studying their development in children. Indeed, if these distances are important, and if the local concentrations of spectral energy they produce are part of the speech task for young children, this should induce adaptive articulatory strategies to cope with the morphological differences. This is the question asked in the present study.

3. Method

Three experiments were conducted. In the first one (section 3.1), natural vowels uttered by 15 speakers from 3 years of age to adulthood were recorded, to describe the acoustic organization of the vowel system during growth. In the second experiment (section 3.2), simulations with an articulatory model were compared with the natural vowels and allowed an interpretation of the articulatory strategies involved in the achievement of acoustic targets. Finally, in the third experiment (section 3.3), the perceptual value of the acoustic targets were investigated using a subset of the natural vowels as stimuli.

3.1. Experiment 1: Production of natural vowels

To study the internal organization of vowel systems during vocal tract growth, we recorded ten isolated occurrences of the French oral vowels /i y u e ø o ɛ œ ɔ a/, produced by speakers at various ages.

3.1.1. Subjects and material

Two groups of children and one group of adults served as subjects for the present study. Each group consisted of five subjects (five females in the 4-year-old group and the

8-year-old group, four females and one male in the adult group). The three groups averaged 3.9 (from 3.7 to 4.2), 8.1 (from 7.9 to 8.3), and 24.6 (from 22.1 to 29.8) years of age. These three groups will be referred to as the 4-year-old group, the 8-year-old group, and the adult group. All subjects were native speakers of Canadian French. The study has been approved by IRB and each subject or their parents signed a consent form before the experiment. The screening procedure consisted of (1) a brief conversation with the experimenter and a speech language pathologist, (2) a 20-dB pure-tone screening at 500, 1000, 2000, 4000, and 8000 Hz, and (3) for children, a brief developmental test to screen for speech production disabilities ("N-EEL: Nouvelles Étude pour l'Examen du Langage", (Chevrie-Muller and Plaza, 2001)). For each speaker, ten repetitions of the ten French oral vowels /i y u e ø o ɛ œ ɔ a/ were elicited in the following forms: "V comme WORD" ("V as in WORD"), where V is one of the ten vowels mentioned above, and WORD is a French word with this vowel in initial position. Only the first vowel V, long and sustained, was analyzed. All speakers repeated the sequence after hearing an adult speaker utter it. The speech signals were recorded in a sound booth with a high-quality tabletop microphone (Sony) at a 15-20 cm distance from the subject's lips, and digitized at 44100 Hz by a *Digital Audio Tape Recorder* (DAT). Signals were then downsampled to 22050 Hz, after low-pass filtering (cut-off frequency of 10000 Hz).

3.1.2. Acoustic analysis

The first three formant frequencies were then extracted for each vowel, using the LPC algorithm integrated in the Praat speech analysis program[5]. The number of poles varied respectively from 10 to 14, in the range of parameters used by Lee *et al.* (1999) and Hillenbrand *et al.* (1995). We used a 14-ms Hamming window, with a pre-emphasis factor of 0.98 (pre-emphasis from 50 Hz for a sampling frequency of 22050 Hz). It is well known that formant measurements are particularly difficult to perform in high-pitched voices, due to the large distance between adjacent harmonics, leading to undersampled spectra. This is especially important for LPC analyses, in which formant measures are greatly influenced by the closest harmonic (Atal and Schroeder, 1974). According to Lindblom (1972), the measurement error could correspond to a value of ± F0/4. However, LPC analysis is the procedure used by recent papers in extensive studies of acoustic characteristics of child speech (Lee *et al.*, 1999; Hillenbrand *et al.*, 1995). Thus, we tried to avoid formant measurement errors by comparing, for each vowel, the automatically extracted formant values overlaid on a wide-band spectrogram with a spectral slice obtained by an FFT analysis with a Hanning window. When important discrepancies were observed either (i) between the overlaid formant values and the spectrogram or (ii) between the overlaid formant values and the spectral slice, the prediction order of the automatic detection algorithm was readjusted and the analysis was performed again. Following Lee *et al.* (1999), in order to evaluate the performance of the LPC analysis on the children's voices, after the automatic analyses were completed, a second experimenter randomly selected 20 tokens and manually measured F1, F2, and F3 on the spectrogram. The differences between the automatically detected formant frequencies and the manually extracted frequencies were the following (in percentage of the mean values): 1.3% (8 Hz) for the first formant, 1.4% (29 Hz) for the second formant and 1.5% (59 Hz) for the third formant. These differences are very small and the measurements can be considered accurate. The formant frequencies were then converted to the Bark scale since this scale models the ear's integration of Hz frequency, following the formula found in Schroeder *et al.* (1979): $F_{Bark} = 7*\text{asinh}(F_{Hz} / 650)$, where F_{Bark} is the frequency, in Bark, and F_{Hz} is the frequency, in Hertz.

3.2. Experiment 2: Modeling vowel production during vocal tract growth

The simulations carried out with the articulatory model aimed at studying the relationships between the articulatory gestures and their acoustic consequences as compared to natural vowels.

3.2.1. The VLAM articulatory model

In this experiment, we used the *Variable Linear Articulatory Model* (VLAM), developed by Shinji Maeda, a scaling of an adult version of Maeda's model (Maeda, 1979) established from cineradiographic data and derived from a statistical analysis guided by knowledge of the physiology of the articulators. VLAM is extensively described elsewhere (Boë, 1999; Ménard *et al.,* 2004). To summarize, VLAM is controlled by seven articulatory parameters (protrusion and labial aperture; movement of the tongue body, dorsum, and tip; jaw height; larynx height) and its main features are the following. It generates 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), and speech signal. Vowels are synthesized by a cascade formant synthesizer excited by a glottal waveform generated by the Liljencrants-Fant source model (Fant *et al.*, 1985; Feng, 1983). The resulting signal is digitized at 22 kHz. 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. In the model, the tongue grows proportionally to the palate. Vocal tract shape can be simulated, month by month and year by year: this was calibrated using the data provided by Goldstein (1980). F0 values are chosen according to the data presented in Beck (1997). VLAM has been compared to real data (Ménard *et al.*, 2004), and it generates realistic articulatory and acoustic vowel configurations. Overall vocal tract lengths and cavity lengths are in line with MRI measurements (Fitch and Giedd, 1999; Vorperian, 2000), and acoustic values obtained for prototypical vowels are in the range of the mean values ± 1 standard error reported for vowels from 3 years old to adulthood (Lee *et al.*, 1999; Hillenbrand *et al.*, 1995). This procedure is thus well-suited for modeling vowel production.

The model was set to 3 growth stages corresponding to the mean age values of the three groups of listeners: a 4-year-old child, an 8-year-old child, and an adult male (21 years old). This age (21 years old) corresponds to the mature stage in the model, when the growth process is terminated (Goldstein, 1980). Fundamental frequency values were set to 300 Hz, 270 Hz, and 110 Hz, respectively for the 4-year-old, 8-year-old, and adult stage. According to Goldstein's data on the ratio of back and front cavity lengths, adult men display a ratio of 1.1, 8-year-olds and 4-year-olds are associated with ratio values of

0.8 and 0.7 respectively. The overall vocal tract lengths obtained were 17.45 cm, 12.65 cm, and 10.67 cm, respectively for the adult male, 8-year-old, and 4-year-old growth stages.

3.2.2. Vowel targets

For each growth stage, 2 sets of targets[6] were generated and compared, based on two criteria: similar articulatory settings to adult or similar acoustic relative positions to adult. First, the maximal vowel space (hereafter MVS), defined as the acoustic space (in the F_1 vs. F_2 and the F_2 vs. F_3 dimensions) generated with a complete coverage of all the possible articulatory values, was simulated (Boë *et al.,* 1989; Ménard and Boë, 2001). A total of about 7000 vowels for each age were retained. This representation (Figure 2) served as a frame of reference for the comparison of generated vowel targets.

[Insert Figure 2 about here]

Then, we situated the four cardinal vowels /i/, /y/, /u/, and /a/, which represent the articulatory-acoustic limits of a speaker, within each of the 3 MVS generated by *VLAM*. Optimal formant triplets (F_1, F_2, and F_3) were determined, based on the following acoustic criteria, inspired from the dispersion-focalization theory (DFT, Lindblom 1996; Schwartz *et al.* 1997):

- [i] : maximal F_3 (focalization of F_3 and F_4), maximal F_2 and minimal F_1
- [y] : F_2 and F_3 close, and minimal F_1 (focalization of F_2 and F_3)
- [u] : minimal F_1 and F_2 (focalization of F_1 and F_2 at the lowest position)
- [a] : maximal F_1 (focalization of F_1 and F_2 at the highest position)

The remaining six French oral vowels were located within that space, based on descriptive studies (Vallée, 1994). For the adult male (21 years old), the underlying articulatory parameters related to each prototypical formant triplet were determined by inversion. Because of the many-to-one relations between articulatory configurations and acoustic targets, many solutions are possible. However, the method exploited here consists in calculating the pseudo-inverse of the Jacobian matrix (Jordan and Rumelhart, 1992). We ensured that the chosen articulatory configuration for each vowel corresponds to earlier descriptive articulatory and acoustic studies of French vowels (Vallée, 1994; Bailly *et al.*, 1995). For the two younger growth stages, the same relative acoustic locations within the MVS were determined for the ten vowels. Again, articulatory parameters were inferred from the same formant-to-articulatory inversion process. These targets, based on similar relative acoustic positions within the MVS throughout growth, will be referred to as “acoustic targets”. They are depicted for the French vowels in the 4-year-old vowel space, in Figure 2.

Then, taking the adult male articulatory-to-acoustic targets as a reference, the values of the seven articulatory parameters related to each French vowel were generated in a 4 and 8 year-old vocal tracts, with non-uniform vocal tract growth. The acoustic values, resulting from unchanged articulatory settings during vocal tract growth, will be referred to as “articulatory targets”.

3.3. Experiment 3: Perception

Three occurrences of each of the ten French oral vowels uttered by each subject were used as a corpus for the perceptual test. The chosen occurrences were the 4th, 5th, and 6th vowels uttered, in order to avoid singular pronunciations (beginning and end-list effects). The duration of the stimuli varied from 242 ms to 307 ms. Each stimulus was presented once, and stimuli of a single speaker were grouped together in a block, followed by a pause. Within each block, stimuli were randomized and blocks were randomized. Each listener heard the stimuli once via high quality headphones, and had to choose the identity of the heard vowel, by clicking on a fixed icon corresponding to the orthographic representation of the vowel on the screen of a computer. Ten choices were available, corresponding to the ten French oral vowels (/i y u e ø o ɛ œ ɔ a/). After the listener had selected the icon, the next stimulus was heard, preceded by a 1-s silence interval. The test took place in a sound booth, and no time constraints were imposed.

20 listeners, aged between 19 and 26, served as subjects for the perceptual test. All were native French speakers from Quebec, enrolled in a speech science degree, or researchers. They did not know, before the test, the goal of the experiment.

4. Results

The production experiment allows a detailed study of the acoustic organization of the vowel space for the three groups of speakers. Simulations with VLAM provide a grid of analysis for the corresponding articulatory-acoustic relationships. The perceptual experiment allows a description of the links between the stable patterns encountered at the production level and the intelligibility of the productions.

4.1. Experiment 1: Produced vowel spaces

4.1.1. Variability

Typical dispersion ellipses (with radii corresponding to ± 1.5 standard error of the individual data around the mean) are given in Figure 3 for 3 speakers representative respectively of the 4-years, 8-years and adult groups, in the F_1 vs. F_2 and in the F_2 vs. F_3 spaces using the perceptual Bark scale for frequencies.

[Insert Figure 3 about here]

First, it can be observed that the acoustic values, for each vowel, differ among the speakers. This phenomenon reflects non-uniform vocal tract growth (Peterson and Barney, 1952; Lee *et al.,* 1999; Hillenbrand *et al.*, 1995). Second, the within-speaker variability, as revealed by the size of the dispersion ellipses, decreases over age, in spite of the Bark correction. The area of each ellipsis, in the F1 vs. F2 space, in Bark, was calculated for each speaker[7]. The vowels were then grouped according to their height, front-backness, and roundedness. Four groups of vowels were considered, along the height dimension: high (/i/, /y/, /u/), mid-high (/e/, /ø/, /o/), mid-low (/ɛ/, /œ/, /ɔ/), and low vowels (/a/). The front vowels were the vowels /i/, /y/, /e/, /ø/, /ɛ/, /œ/, and the back vowels consisted of /u/, /o/, and /ɔ/. In the front series, /i/, /e/, and /ɛ/ were the unrounded vowels and their rounded counterparts were /y/, /ø/, /œ/. Note that since the phoneme /a/ is sometimes realized as a front low vowel [a] and a back low vowel [ɑ] in French, we did not specify this vowel with respect to front-backness and roundedness. For instance, the mean dispersion of the high vowels for the 4-year-old speakers consists of the mean area of the vowels /i/, /y/, and /u/, for these speakers. These values are given for each speaker group in Figure 4. Mixed ANOVAs carried out on the area values, as the dependant variable, and the feature and age group as the factors, reveal a developmental pattern. The effect of age is indeed significant (height: $F(2, 153)=16.82$, $p<0.05$; front-backness: $F(2, 159)=24.57$, $p<0.05$; rounding: $F(2, 159)=14.67$, $p<0.05$). Planned comparisons further reveal that for all three features, the areas of the dispersion ellipses are larger for the 4-year-old speakers, on the one hand, compared to the 8-year-old and adult speakers, on the other hand (height: $F(1, 153)=36.29$, $p<0.05$; front-backness: $F(1, 159)=47.09$, $p<0.05$; roundedness: $F(1, 159)=27.52$, $p<0.05$). The height factor does not have a significant effect on the area of the dispersion ellipses, neither as a main effect nor in interaction with the age factor. Similar results are found for the rounding factor. However, the analyses reveal that the front-backness factor has a significant effect on the area ($F(1, 159)=19.605$, $p<0.05$), back vowels being associated with larger ellipses than front vowels. This pattern can be explained by the more crowded space in the front area, where the rounding distinction is realized. The interaction between front-backness and age is also significant ($F(2,159)=3.7662$, $p<0.05$), revealing that the difference in area is more important for 4 and 8-year-old speakers, compared to adult speakers. These results are in line with previous work which demonstrates that increased acoustic variability, typical of young speakers, reflects immature motor control abilities (Eguchi and Hirsh, 1969; Smith and Goffman, 1998).

[Insert Figure 4 about here]

Despite this important between-speaker and within-speaker variability, it is striking to observe some regularity in the organization of the vowel systems. The next sections focus on the extreme vowels /i/, /u/, /a/, and /y/, which define the limits of the F1 vs. F2 and F2 vs. F3 vowel space.

4.1.2. Stable patterns associated with the vowels /i/, /u/, /a/, and /y/

Even though the internal organization of the vowel systems (in terms of vowel location and size of the dispersion ellipses) varies across speakers, the four cardinal vowels /i/, /u/, /a/, and /y/ are always produced at the extreme limits of the F1 vs. F2 vs. F3 acoustic space. Indeed, /i/ defines the upper left corner of the space (lowest F1, highest F3), /u/ corresponds to the upper right corner (lowest F1, lowest F2 and low F3), /a/ is produced at the lower limit of this space (highest F1), and /y/ is located at the lower left corner of the F2 vs. F3 space (lowest F1). As a result, the spectral patterns of these vowels conform to the acoustic targets shown in Figure 2, based on criteria of maximal distance and local stability. The latter is related to a reduced distance between two formants, which creates a strong concentration of spectral energy, or "focalization" (Abry *et al.,* 1989). For /i/, this concentration of spectral energy is created by the proximity of F3 and F4, whereas it corresponds to the proximity of F1 and F2 for /u/. As regards /a/, focalization is related to the proximity of F1 and F2, where F1 is maximal. The vowel /y/ shows close F2 and F3, with minimal F1 values. However, according to the predictions of section 2 (Figure 1), we should observe some variation in these formant distances, if the three groups of speakers were using the same articulatory strategies. The next section attempts to use the simulations carried out with the articulatory model to determine whether similar or different articulatory strategies are required to reach a focal target.

4.2. Experiment 2: Simulations with an articulatory model

Results of experiment 1 showed that the vowels /i/, /y/, /a/, and /u/ are produced by the 4-year-old, the 8-year-old, and the adult groups using a strong concentration of spectral energy related to close formants (focalization). The present section is devoted to an investigation of articulatory configurations related to focal vowels across ages.

4.2.1. Focalization

Figure 5 represents, for the 4-year-old MVS, the acoustic targets (stars) and the articulatory targets from the adult male MVS (circles) in VLAM. Acoustic targets are presented for the ten French oral vowels, whereas articulatory targets are presented only for the four vowels /i/, /y/, /u/, and /a/. For the sake of clarity, the articulatory and the acoustic targets for a given vowel are linked by a solid line.

[Insert Figure 5 about here]

As described in section 3.2.2, articulatiry targets represent the acoustic results of adult articulatory strategies for this vowel, generated in a 4-year-old vocal tract, whereas acoustic targets correspond to relative acoustic locations, defined by the DFT (dispersion-focalization theory i.e. maximal contrast and focalization). Acoustic targets are thus related to different articulatory strategies for the child, compared to the adult. The distance between each target, for /i/, /y/, /u/, and /a/, can be seen as the acoustic difference between an articulatory-based target (articulatory targets) and an acoustic-based target (acoustic targets). Such a difference is very small for /u/ and /a/. However, an important distance between the acoustic and the articulatory targets can be found for /i/ (in the F2 vs. F3 space) and /y/ (in the F1 vs. F2 and the F2 vs. F3 spaces).

In order to compare the acoustic consequences of different articulatory gestures, we calculated the F1-F2 distance, for /u/ and /a/, the F3-F4 distance for /i/, and the F2-F3 distance for /y/, for the synthesized vowels and the natural vowels described in the previous section. Results are shown in Figure 6. All data are in Bark. Note that the absolute values of formant differences vary across vowels, for anatomical reasons (Fant, 1975; Boë, 1999). The dashed line corresponds to the acoustic targets, the dotted line stands for the articulatory targets, and the solid line corresponds to the values produced by our speakers. The lower the y-value, the more focal the vowel. One-way ANOVAs were carried out for each of the four vowels, considering the value of distance for the natural vowels, in Bark, as the dependent variable, and the speaker's age as the factor. No significant difference was observed for /i/, /y/, and /a/. As concerns /u/, the ANOVA revealed a statistical difference between the 4 year-old group and the two other groups ($F(1,12)=12.62$, $p<0.05$). Surprisingly, the younger group is producing even more focal /u/ than the 8-year-old and the adult speakers. Thus, the formant distances for /i/, /u/, /a/, and /y/, giving rise to focal vowels, for the 4-year-olds and the 8-year-olds are globally comparable to those of the adult speakers. This pattern is important considering the large differences in the size of the cavities between the children and the adults, as discussed previously.

[Insert Figure 6 about here]

4.2.2. Adaptative articulatory strategies

Comparing now the values produced by the speakers to the articulatory and acoustic targets modeled from VLAM, it can be seen that the distance values in the vowels /i/, /u/, and /a/ synthesized by the acoustic and the articulatory targets do not vary much from 4 year-old to the adult stage (21 years old, in the model). The difference is indeed of 1 Bark or less. However, the difference in distance values between articulatory and acoustic targets for the vowel /y/ is greater and decreases with increasing age. A difference of more than 2 Barks can be observed at 4 years of age. In that case, it is striking to see that the values produced by the speakers closely conform to the values associated with the acoustic targets. Since focalization was part of the goal related to these vowels, we suggest that even at 4 years of age, young speakers adapt their articulatory strategies to produce a focal acoustic target. Basically, this means that the young speakers, taking the adult male as a reference, should exploit a more anterior position of the tongue body, thus shortening the front cavity, in order to increase its affiliated formant - that is F_2 for /y/ and F3 for /i/.

4.3. Experiment 3: Perceptual results

The acoustic variability encountered in the vowel systems produced by children and adults may keep them from reaching the correct perceptual targets associated with the French oral vowels. Furthermore, the focalization of the spectral patterns associated with the vowels /i/, /y/, /u/, and /a/ may require different articulatory gestures for children and adults, to cope with anatomical differences. The current section presents the results of the perceptual test aimed at describing the intelligibility of the produced vowels.

4.3.1. Global intelligibility

The confusion matrices for the three groups of speakers are presented in Table 1, in percentage of the maximal number of responses (300 for a given vowel). The intelligibility scores for each vowel correspond to the bold shaded cells of the matrices. First, it can be observed that the vowel associated with the lowest intelligibility is the mid-low and back vowel /ɔ/ (58.3% for the adults, 59% for the 8-year-olds, and 67.7% for the 4-year-olds). This perceptual confusion can however be ascribed to a phonological neutralization occurring in Canadian French, by which this phoneme category merges with the low vowel /a/ (Santerre, 1976). In order to better represent the global intelligibility, we calculated the mean score for each speaker, without considering the phoneme /ɔ/. Figure 7 plots the data, with the standard errors. ANOVAs show that the age factor has a significant effect on the global intelligibility scores ($F(2, 12)=4.2763$, $p<0.05$). The score associated with the 4-year-old group is well below the score of the 8-year-old and adult groups.

[Insert Figure 7 about here]

[Insert Table 1 about here]

4.3.2. Feature intelligibility

Intelligibility scores were then calculated according to the three features of height, front-backness, and roundedness. The scores are displayed in Figure 8 for the three features. Concerning openness (upper left graph in Figure 8), the number of correct responses increases from the 4-year-old group to the 8-year-old group for high (/i/, /y/, /u/) and mid-high (/e/, /ø/, /o/) vowels. However, a one-way ANOVA carried out on feature intelligibility scores with age as the factor shows that the tendency reaches significance only for the values corresponding to high vowels ($F(2,12)=3.86$, $p<0.05$). Based on the results of planned comparisons, no difference can be observed between the 8-year-old and the adult groups whereas the 4-year-old group is significantly different from the 8-year-old and the adult speakers pooled together ($F(1,12)=7.67$, $p=0.05$). Note that the low vowel /a/ is associated with a nearly perfect identification score for all three groups of speakers. The intelligibility of the mid-low vowels (/ɛ/, /œ/, /ɔ/) does not increase with the age of the speaker. This group is the least intelligible one for the 8-year-old and the adult speakers. The phonological neutralization of the back mid-low vowel /ɔ/ in Canadian French, suggested above, is possibly responsible for this lower intelligibility. As for the front-backness dimension (upper right panel in Figure 8), for the three speaker groups, the front vowels are more intelligible than the back vowels. Intelligibility for the front vowels increases with speaker age. This tendency is significant ($F(2,12)=4.30$, $p<0.05$), with the 4-year-old group having lower scores than the 8-year-old and the adult groups ($F(1,12)=8.50$, $p<0.05$). As concerns the rounding contrast, as can be seen on the lower graph of Figure 8, the 4-year-old group is associated with the lowest percentage of correct answers. This tendency is only significant for the unrounded vowels ($F(2,12=4.60$, $p<0.05$), with the 4-year-old group having lower scores than the 8-year-old and adult groups ($F(1,12)=9.19$, $p<0.05$).

[Insert Figure 8 about here]

In order to summarize the data, Figure 9 provides a schematic representation of the main confusions identified from the perceptual results, for the adults and the 8-year-olds, on the one hand, and on the other hand, for the 4-year-olds. Those are the vowels associated with a score lower than 85% (/i/, /y/, /u/, /e/, /ɔ/). It can be seen that the previously discussed neutralization between /ɔ/ and /a/ is present for all groups. More importantly, lowest intelligibility scores for the 4-year-olds mainly concerned high vowels: /i/ is perceived /e/, /y/ is perceived /i/ and /u/ is perceived /o/. The confusion between /e/ and /ɛ/ is surprising, and may be ascribed to the late development of this specific phonological contrast.

[Insert Figure 9 about here]

5. Discussion

Findings from this study (experiment 1) show that the internal organization of the vowel spaces produced by 4-year-old, 8-year-old, and adult speakers can be characterized by an important variability. However, speakers from the three age groups tend to produce focal vowels /i/, /y/, /u/, and /a/, which delimit the F1 vs. F2 vs. F3 vowel space. Simulations with an articulatory model lead us to consider more specifically the vowel /y/, for which the production of such patterns likely involves different articulatory strategies for the 4-year-olds compared to the adults. These results suggest a recalibration of articulatory-to-acoustic relationships during growth, in order to achieve focalization. This goal is important even though it prevents the child from achieving well perceived /y/. Implications for the "dispersion-focalization" theory of vowel systems, in the framework of a perception for action control theory (PACT, Schwartz *et al.*, 2002), are presented in section 5.3.

5.1. Variability and stability: the importance of focalization

Our acoustic analyses of natural vowels revealed that the size of the dispersion ellipses, in the F1 vs. F2 space (in Bark) decreases with age. Indeed, along all three features (height, front-backness, and roundedness), ellipses are significantly broader for 4-year-olds, compared to 8-year-olds, and adults. These broader ellipses are likely due to increased variability which has been interpreted as reflective of immature motor control (Eguchi and Hirsh, 1969; Smith and Goffman, 1998; Green *et al.*, 2000). In our acoustic spaces, such immaturity gives rise to larger dispersions around the mean, for a given vowel target.

This important variability is also accompanied by a remarkably stable pattern for the vowels /i/, /y/, /u/, and /a/, namely focalization, defined as a strong concentration of spectral energy related to close neighboring formants. For /i/, the distance between F3 and F4 is small, whereas for /y/, F2 and F3 are close together. Focalization, for /u/ and /a/, is related to a reduced distance between F1 and F2, F1 being at the lowest position for

/u/, and at the highest position for /a/. This stable pattern is striking considering the predicted effects of non-uniform vocal tract growth on the resulting formant patterns. As schematized in Figure 1, since the pharyngeal cavity is much shorter than the oral cavity in the young child whereas the pharyngeal cavity is longer than the oral cavity for the adult male, formants affiliated with the back cavity should decrease much more than formants affiliated to the front cavity during growth. As a result, assuming similar articulatory positions for the 4-year-old and the adult, the distance between F3 and F4 for /i/, F2 and F3 for /y/, and F1 and F2 for /a/ and /u/ should be larger in the child compared to the adult. Comparable formant distances suggest the use of adaptive articulatory strategies.

In order to compare the acoustic results of similar acoustic focal targets (referred to as acoustic targets) and similar articulatory strategies (referred to as articulatory targets), simulations with an articulatory model were conducted. A comparison of formant distances corresponding to these synthesized vowels and natural vowels shows that speakers tend to produce formant distances that conform to the acoustic targets (Figure 6). This tendency is particularly striking for the vowel /y/, for which the difference between F2 and F3 is much reduced for the acoustic targets compared to the articulatory targets.

5.2. Focalization at the detriment of intelligibility

5.2.1. The case of /y/

Speakers as young as 4 years of age tend to produce the vowel /y/ with a reduced distance between F2 and F3. According to the simulations presented in section 4, producing such a focal /y/ involves the exploitation of different articulatory strategies for children and adult males, to cope with the relatively smaller pharynx of younger speakers. The primacy of acoustic targets during vocal tract growth, defined following focalization criteria inspired from the DFT (section 3), is in line with Stevens' quantal theory (Stevens, 1989), in which formant convergence is related to articulatory-to-acoustic stability.

How can this acoustic target and the involved adaptive articulatory gestures be related to a lower intelligibility for the child? Considering /y/, an examination of vocal tract morphology and formant-cavity affiliations can be useful. Because of the overall shorter vocal tract, the resonant frequencies of the pharynx and the oral cavity are higher for the child compared to the adult male. We have shown in a previous study (Ménard *et al.,* 2002) that F_2', corresponding to a weighted sum of F2, F3, and F4, represents a good predictor of perceived roundedness. A perceived rounded vowel corresponds to an F_2' value below 15 Bark, whereas F_2' of a perceived unrounded vowel is above 15 Bark. For the adult, reaching an F_2' value lower than 15 Bark is possible together with a focalization of F_2 and F_3: the long cavities ensure low F2 and F3 values, and the F_2 and F_3 complex is sufficiently low to be below 15 Bark. On the contrary, for the young child, F2 and F3 are high. The "rounding target" with F_2' lower than 15 Bark is achievable if no fronting or no compensation strategy is used (this case is displayed in Figure 1 and corresponds to the articulatory target). In this case, F2 and F3 are too far apart to merge, and F_2' corresponds to F_2, hence it is set at a value of 13.96 Bark, well below the 15-Bark limit. However, realizing a focal vowel (acoustic target) involves close F_2 and F_3, which increases this whole group to a value above (or at the limit of) the 15-Bark category boundary. Actually, F_2' for the focal acoustic target is set at 16.15 Bark. These competing constraints would lead to a decrease of intelligibility for the 4-year-old group, the vowel /y/ being often perceived /i/.

Hence, it seems that focalization plays such a crucial part in shaping the speaker's vowel system during growth that it is achieved at the cost of lower intelligibility for young children. With similar perceptual constraints (F_2'<15 Bark), one can consider that vocal tract shapes representative of very young speakers sometimes prevent them from achieving intelligible /y/ vowels in French, focalization being their major goal. The importance of focalization in the production domain is also found in the perception domain. Indeed, Polka and Bohn (2003) report a robust asymmetry effect in discrimination experiments, in adults as well as in children. They interpret this effect as a preference for peripheral vowels, providing an anchor for comparison. In the framework of the DFT (Schwartz *et al.*, 1997), this pattern can be interpreted as a preference for focal vowels: these vowels are preferred perceptually because of their perceptual salience provided by two close formants (Schwartz *et al.*, 2005). Notice that the rounded/protruded shape of the lips should provide a clear visual correlate of the rounded feature, compensating to a certain extent the lack of auditory intelligibility.

5.2.2. The vowels /i/ and /u/

Two other focal vowels were associated with a lower intelligibility score with decreasing age, namely /i/ and /u/, respectively perceived /e/ and /o/. These vowels are thus perceived as being more open, which is similar for the mid vowel /e/, also misidentified /ɛ/. The confusion between height degrees can be understood in the light of a non-linear decrease in formants and F0 during vocal tract growth. We have shown elsewhere that the difference between F1 and F0, in Bark, is a good acoustic predictor of perceived height in French (Ménard *et al.*, 2002). Perceived high vowels are associated with F1-F0 lower than 2 Bark, perceived mid-high vowels, to F1-F0 greater than 2 Bark and lower than 4 Bark, and perceived mid-low and low vowels are related to F1-F0 greater than 4 Bark. As concerns high vowels, F1 is affiliated to a Helmholtz resonator and thus remains relatively low during growth (Lee *et al.*, 1999; Ménard *et al.*, 2004). However, the mean F0 value decreases from 436 Hz at birth to 112 Hz at adulthood (Beck, 1997), a difference of about 3 Bark. Thus, once the appropriate articulatory strategies for controlling Helmholtz resonators are mastered, low F1 values combined with high F0 values would both contribute to produce a small F1-F0 difference, related to perceived high vowels. However, during growth, decreasing F0 combined with keeping F1 relatively constant increases the value of F1-F0. High vowels would no longer be perceived as high, but as mid-high, since F1-F0 would be greater than 2 Bark, the category boundary between high and mid-high vowels. Unlike /y/, focalization does not contribute to the low intelligibility of these vowels for young speakers. Rather, non-linear transformations of the laryngeal and the supralaryngeal cavities would be related to perception errors. The effects of such non-linearity are currently under study.

5.3. Results in the light of dispersion, focalization
and the perception for action control theory (PACT)

Schwartz *et al.* (2002) proposed a theory of speech motor control, called the perception for action control theory (PACT), based on the ability of the perceptual system to recover phonological goals from an incoming signal, and to guide the speaker's production system. The data presented above raise interesting questions that can be reinterpreted in the light of dispersion, focalization and the PACT.

According to Locke (1983), the young speaker builds and uses the perceptual templates of his/her language in order to guide the speech task during vocal tract growth. Through imitation, articulatory settings related to an acoustic product that conforms to the perceptual template defining the intended phonological entity are internalized. Our assumption is that the child needs to recalibrate this articulatory-to-acoustic mapping, as the overall vocal tract length increases and as the ratio of the pharynx length versus the oral cavity length increases. The remarkable aspect of the present results is that this overall recalibration seems to follow a pattern globally compatible with the Dispersion-Focalization Theory, a theory that had been primarily proposed for dealing with phylogeny rather than ontogeny (Schwartz *et al.,* 1997). Indeed, children and adults seem to organize their production in terms of sufficient dispersion (Lindblom, 1986) and focalization in the acoustic-perceptual space. Regarding rounding, a low F_2' (lower than 15 Bark), F_2-F_3 focalization and lip protrusion seem to be part of the /y/ target, which results in the tongue body having to be more fronted for the child, compared to the adult. This provides a very strong indication that focalization is indeed part of the perceptual goal for speech production (Abry *et al.,* 1989). We propose that this goal is related to the fact that, just as /i/, /u/, and /a/ define the limit of the F1- F_2 vowel space, /i/, /y/, and /u/ define the extreme limits of the F_2 and F_3 space.

Producing /y/ is in that case defining one limit of this space. The perceptual system recovers this pattern in adult speech, and uses it to control the speaker's production. The guiding template is a spectral prominence pattern, defined by focalization of F_2 and F_3, below 15 Bark.

6. Conclusion

The aim of the present paper was to describe some production-perception relationships observed in the realization of French oral vowel systems, during non-uniform vocal tract growth. Recordings of real data showed that focalization (a spectral pattern involving close adjacent formants) is produced by the speakers, and hence seems to be part of their goal. Simulations with the VLAM articulatory-to-acoustic model revealed that for young children, an adaptive articulatory strategy is required in order to reach focalization, namely a fronting of the tongue body. This pattern, however, results in the production of less intelligible vowels /y/, for the 4-year-old speaker. This feature is thus realized at the cost of intelligibility. These results have important implications in the framework of the production-perception relationships during growth, and suggest further investigation in the field of speech motor control development.

Acknowledgements

This work was supported by a grant from the Social Sciences and Humanities Research Council of Canada, and by a grant from the Fonds Québécois de la Recherche sur la Société et la Culture. We would like to thank Dr Amanda Walley and three anonymous reviewers for insightful comments and suggestions. We are indebted to Shinji Maeda and Marija Tabain for fruitful discussions.

References

Abry, C. & Boë, L.-J. (1986). "Laws" for lips. *Speech Communication*, 5, 97-104.

Abry, C., Boë, L.-J. & Schwartz, J.-L. (1989). Plateaus, catastrophes and the structuring of vowel systems. *Journal of Phonetics*, 17, 47-54.

Atal, B. S. & Schroeder, M. R. (1974). Recent advances in predictive coding – applications to speech synthesis. in Fant, G. (ed.), *Speech Communication*, Almqvist and Wiksell, Sweden, 1, 27-31.

Badin, P., Perrier, P., Boë, L.-J. & Abry, C. (1990). Vocalic nomograms : Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America*, 87, 1290-1300.

Bailly, G., Boë, L.-J., Vallée, N. & Badin, P. (1995). Articulatory-acoustic vowels prototypes for speech production. *4th European Conference on Speech Communication and Technology*, Madrid, 3, 1913-1916.

Beck, J. M. (1997). Organic variation of the vocal apparatus. in Hardcastle, W. J. et Laver, J. (eds), *Handbook of Phonetic Sciences*, 256-297.

Boë, L.-J., Perrier, P., Guérin, B. & Schwartz, J.-L. (1989). Maximal Vowel Space. *1st European Conference on Speech Communication and Technology (Eurospeech)*, 2, 281-284.

Boë, L.-J. (1999). Modeling the growth of the vocal tract vowel spaces of newly-born infants and adults. Consequences for ontogenesis and phylogenesis. *Proceedings of the International Congress of Phonetic Sciences*, 3, San Francisco, 2501-2504.

Chevrie-Muller, C. & Plaza, M. (2001). *Nouvelles Étude pour l'Examen du Langage*, Paris, Ecole de Psychologie Appliquée.

Chiba, T. & Kajiyama, M. (1941). *The vowel : its nature and structure*, Tokyo-Kaiseikan.

Eguchi, S. & Hirsch, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngol.* Suppl. 257, 5-51.

Fant, G. (1960). *Acoustic Theory of Speech Production*, Mouton, Le Hague.

Fant, G. (1975). Non-uniform vowel normalization. *KTH, Speech Transmission Laboratory-Quarterly Progress and Status Report*, 2-3, 1-19.

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

Fant, G. & Båvegård (1997). Parametric model of the vocal tract area function: Vowels and consonants. *TMH-QPSR*, 1, 1-20.

Feng, G. (1983). Vers une synthèse par la méthode des pôles et des zéros. *Actes des Journées d'Etude sur la Parole*, Groupe Francophone de la Communication Parlée, SFA, 155-157.

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 growing children*, Doctor of Science thesis, MIT, Cambridge, Massachusett.

Green, J. R., Moore, C. A., Higashikawa, M. & Steeve, R. W. (2000). The Physiologic Development of Speech Motor Control : Lip and Jaw Coordination. *Journal of Speech, Language and Hearing Research*, 43, 239-255.

Hillengrand, 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.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward Models: Supervised Learning with a Distal Teacher. *Cognitive Sciences*, 16, 316-354.

Kent, R. D. (1976). Anatomical and Neuromuscular Maturation of the Speech Mechanism : Evidence from Acoustic Studies. *Journal of Speech and Hearing Research*, 19, 421-447.

Kuhl, P. K. (1992). Speech prototypes : Studies on the nature, functions, ontogeny and phylogeny of the « centers » of speech categories. in Tohkura, Y., Vatikiotis-Bateson, E. et Sagisaka, Y. (eds), *Speech perception, production and linguistic structure*, Tokyo, Ohmsha, 239-264.

Lacerda, F. (2003). Phonology: And emergent consequence of memory constraints and sensory input. *Reading and Writing: An Interdisciplinary Journal*, 16, 41-59.

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.

Lindblom, B. (1972). Comments on paper 15 'Development of speech sounds in children' by Eguchi, S. and Hirsh, I. J. . in Fant, G. (ed), *International Symposium* *on Speech.*

Lindblom, B. (1986). Phonetic universals in vowel systems. in J.J. Ohala and J.J. Jaeger (eds), *Experimental Phonology*, New-York, Academic Press, 13-44.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America* 99, 1683-1692.

Locke, J. L. (1983). *Phonological Acquisition and Change*, New-York, Academic Press.

MacNeilage, P. F. & Davis, B. L. (2000). On the Origin of Internal Structure of Word Forms. *Science*, 288, 527-531.

MacNeilage, P. F. (1997). Acquisition of speech. in W. J. Hardcastle and J. Laver (eds), *Handbook of Phonetic Sciences*, Cambridge, Blackwell Publishing, 301-332.

Maeda, S. (1979). An Articulatory Model of the Tongue Based on a Statistical analysis. *Journal of the Acoustical Society of America*, 65, S22.

Ménard, L. & Boë, L.-J. (2000). Exploring Vowel Production Strategies from Infant to Adult by Means of Articulatory Inversion of Formant Data. *International Congress of Spoken Language Processing*, Beijing, China, 465-468.

Ménard, L., & Boë, L.-J. (2001). "Perceptual categorization of maximal vowel space from birth to adulthood," *European Conference on Speech Communication and Technology (Eurospeech)*, Aalborg, Danemark, 167-170.

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S. & Vallée, N. (2002). Auditory normalization of french vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America*, 111, 4, 1892-1905.

Ménard, L., Schwartz, J.-L. & Boë, 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.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 351-382.

Nordström, P.-E. (1975). Attempts to simulate female and infant vocal tracts from male area functions. *KTH, Speech Transmission Laboratory-Quarterly Progress and Status Report*, 2-3, 20-33.

Peterson, G. E. & Barney, H. L. (1952). Control method used in the study of vowels.

Journal of the Acoustical Society of America, 24, 175-184.

Polka, L. & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221-231.

Santerre, L. (1976a). *Fréquence des phonèmes en français montréalais, dans 184 000 occurences du corpus Sankoff*. Départ. linguistique, Université de Montréal.

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 Lindblom, B. and Öhman, S. (eds), *Frontiers of Speech Communication Research*, London, Academic, 217-229.

Schwartz, J.L., Abry, C., Boë, L.J. & Cathiard, M. (2002). Phonology in a theory of perception-for-action-control. In J. Durand, B. Laks (eds.) *Phonetics, Phonology and Cognition* (pp. 255-280). Oxford: Oxford University Press.

Schwartz, J.-L., Beautemps, D., Abry, C. & Escudier, P. (1993). Inter-individual and cross-linguistic strategies for the production of the [i] vs. [y] contrast. *Journal of Phonetics*, 21, 411-425.

Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997). .The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, 25, 255-286.

Schwartz, J.L., Abry, C., Boë, L.J., Ménard, L. & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication*, 45, 425-434.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.

Vallée, N. (1994). *Systèmes vocaliques : de la typologie aux prédictions*, Ph. D. Thesis, Grenoble, Université Stendhal/Institut de la communication parlée.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R. & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America*, 117, 338-350.

Vorperian, H. K. (2000). *Anatomic development of the vocal tract structures as visualized by MRI*. Ph. D. dissertation, Madison, University of Wisconsin-Madison.

[1] We will refer to the cavity between the constriction location and the lips by the term « front cavity », whereas « back cavity » refers to the cavity behind the constriction location. The terms « oral » and « pharyngeal » are used to refer to the physiological cavities from the pharyngeal wall to the lips, and from the velum to the larynx, respectively.
[2] The front cavity resonance for /y/ can also be considered as a Helmholtz resonance considering lip closure (Badin *et al.,* 1990). This does not change the reasoning in the following sections.
[3] Note that in French, formant cavity affiliations for F1 and F2 can also be reversed, F2 being affiliated with the Helmholtz resonance of the front and labial cavities, and F1 being affiliated to the Helmholtz resonator created by the constriction and the pharyngeal cavity.
[4] The ratio of the cavity lengths is not the only factor for explaining the F2 vs. F3 pattern. Besides these resonances, laryngeal tube resonance affects F3 in high vowels (Chiba and Kajiyama, 1941; Fant and Båvegård, 1997).
[5] Available at http://www.fon.hum.uva.nl/praat/
[6] We define as a "target" a point of reference (be it acoustic or articulatory) for the articulatory model, which remains unchanged during vocal tract growth. We do not consider this landmark as a perceptual target.
[7] If a and b are respectively the lengths in Bark of the major and minor axes of the dispersion ellipsis, provided by the diagonalisation of the covariance matrix, the area is provided by: Area=π a b.