Acoustic variability and adaptive articulatory strategies during vocal tract growth revealed by the rounding contrast in French

Lucie Ménard
Université du Québec à Montréal, Canada
E-mail: menard.lucie@uqam.ca

ABSTRACT

This paper reports on the articulatory and acoustic variability involved in the realization of the rounding contrast in French during vocal tract growth. Data are taken from a database of ten French vowels uttered by 12 speakers ranging in age from 4 years to adulthood. Despite the important acoustic variation, it appears that focalization (close F2 and F3) for /y/ is a pattern well realized by all the speakers. Our simulations using an articulatory model demonstrate that the realization of this feature involves a more anterior position of the tongue body for young children, an adaptive articulatory strategy exploited to compensate for their small pharynx. Perceptual tests, however, reveal that this focalization prevents the young child from achieving the perfect perceptual objective related to the rounded vowel /y/. We argue that this is due to the peripheral constraints associated with young children's vocal tract configuration.

1. INTRODUCTION

Non uniform vocal tract growth has been an object of study for decades. In the search for normalization factors, some studies aimed at determining scaling factors that could explain acoustic variation observed for vowels produced by men, women, and children. Most of the time, even though a precise acoustic or geometrical modelling is attempted, cavity length alone can not explain all the acoustic variation encountered.

As a result of articulatory modelling, the acoustic effects of non uniform vocal tract growth can be predicted and compared to observed variability. Thus, the general trends associated with non uniform growth of the cavities (oral 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 modelling, 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/ ([1]). 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?

2. THE ROUNDING CONTRAST DURING VOCAL TRACT GROWTH

2.1. Global predictions

The articulatory gestures underlying the /i/ versus /y/ contrast in French for adult speakers are well known ([2]). 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. Since the constriction area is in this case very small (about 0.3 cm^2), the front and back cavities act as simple tube resonators and their resonant frequency is very similar to the formants. Coupling effects are thus minimal and formant-cavity affiliations can be considered. The configuration created by the whole vocal tract corresponds to a Helmholtz resonator and is affiliated to F1. F2 is the half wavelength resonance of the back cavity (R2) and F3, the half wavelength resonance of the front cavity (R3)

Compared to /i/, the basic gesture associated with the vowel /y/ is a 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. If the front cavity remains shorter than the back cavity, F2 is affiliated to the half wavelength resonance of the back cavity (R2) and F3, to the half wavelength resonance of the front cavity (R3). On the contrary, 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, F2 is affiliated to the front cavity (R3) and F3, to the back cavity (R2). The acoustic consequences of these gestures are described in detail in [3].

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 much longer than the oral cavity ([4]). This non uniform growth of the two cavities has important effects on the formant patterns of vowels, since the shorter the cavity, the higher the formant. Considering the two vowels /i/ and /y/ described earlier, all the articulatory gestures remaining unchanged, the following predictions can be made, if we consider the adult male values as a reference. First, the formant affiliated to the back cavity, including the pharynx, will be much higher than the formant affiliated to the front cavity, for the infant. Thus, the differences between R2 and R3 will be directly related

 ISBN 1-876346-48-5

to the ratio of the length of the pharyngeal cavity to the length of the oral cavity: the smaller the ratio, the smaller the difference between R2 and R3 for /i/. In the case of /y/, for the infants, protrusion of the lips will result in a front cavity much longer than the back cavity. As a result, R3 will be much lower for the infant, compared to the adult male. Contrastingly, for /y/, the smaller the ratio, the larger the difference between R2 and R3.

2.2. Subjects and material

In order to compare these global predictions to natural vowels, 12 speakers were recorded, from 4 years of age to adulthood, by steps of 2 years. Sexual differences were taken into account after 10 years of age, based on previous data ([5]). As a consequence, one speaker was representative of the ages 4, 6, 8, and 10, and two speakers (a male and a female) were recorded for each of the ages 12, 14, 16, and adulthood. For each speaker, the ten French oral vowels /i y u e ø o ɛ œ ɔ a/ were ellicited by the following forms: “i comme hibou” (*i as in hibou*). Only the first vowel, long and sustained, was analysed. At least ten repetitions were obtained. Speakers from 4 years old to 12 years old repeated the phrase after hearing an adult French speaker utter it, while older speakers read directly from the corpus. All subjects were native French speakers from France and had no known speech production or hearing disability. The speech signals were recorded in a sound-treated booth, and digitized at 44100 Hz onto DAT recorder, via a high quality microphone.

Formant frequencies at the middle of the vowel duration were then extracted for each vowel, using the LPC algorithm and the Burg formant extraction algorithm, both integrated in the Praat speech analysis program. For all the vowels, the results of the analysis were compared to formant values based on a visual inspection of the spectrographic representation. In case of important discrepancies between the two methods, parameters for the automatic detection algorithm were readjusted and the analysis was performed again.

2.3. Focalization

Figure 1 represents the dispersion ellipses (±1.5 s. d. to the mean) of the ten occurrences of each French vowel, for three speakers: the 4 year-old child (ch_4_f), the adult male (fb_23_m), and the adult female (mc_39_f). It is noticeable that for the three speakers, /y/ is a focal vowel, that is, a vowel with a strong concentration of spectral energy ([2]) realized by a formant convergence (F2 and F3). The same pattern is realized by all the speakers but one (the 12-year-old female). All the formant values were converted to Bark units , and the difference between F2 and F3 were calculated. The resulting differences were lower than 2 Barks (except for the 12-year-old female speaker), a value well in line with results gathered from adult French speakers ([3]). This pattern is particularly striking if one considers, based on the predictions of section 2.1, that a vowel /y/, articulated with similar articulatory positions but in vocal tracts representative of an adult male and a young child, would correspond to quite different formant patterns. For the former, F2 and F3 would be close (focal), whereas for the latter, F2 and F3 would be far appart, so that focalization is not achieved.

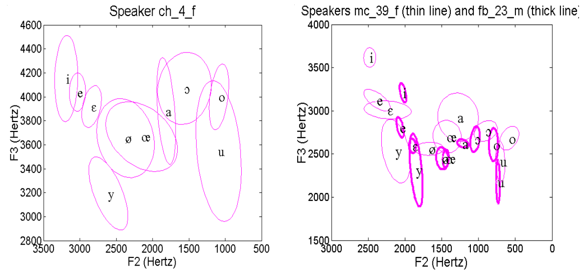

Figure 1: Dispersion ellipses, in the F2 vs F3 space, of the ten French oral vowels, for the 4-year-old speaker (left panel) and the adult speakers (right panel).

3. VOWEL INTELLIGIBILITY

The analysis of natural vowels showed that focalization is well realized by almost all the speakers. Is this acoustic pattern achieved in order to realize intelligible vowels? To answer this question, a perceptual test was carried out.

3.1. Perceptual tests

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). 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 stimulus once via high quality headphones, and had to choose, by clicking on an icon on the screen of a computer, the identity of the heard vowel. Ten choices were available, corresponding to the ten French oral vowels. The test took place in a sound-treated booth, and no time constraints were imposed. 20 listeners, aged between 19 and 43 years, served as subjects for the perceptual test. All were native French speakers from France, enrolled in a speech science degree, or researchers of our laboratory. They did not know, before the test, the goal of the experiment.

The results of the test were considered, first, in terms of the number of erroneous identifications by adult listeners. Strikingly enough, a developmental pattern was noticeable in the perception of the rounding contrast: the rounded vowels /y/ intended by the 4-year-old speaker were most frequently perceived as unrounded /i/. For older speakers, produced /y/ were correctly identified. Thus, when achieving the task related to these vowels, our 4-year-old speaker often misses the perceptual target, producing vowels perceived as unrounded. Such a developmental pattern did not appear for perceived height and place of articulation.

3.2. Acoustic correlates of perceived rounding

In order to describe the low intelligibility of /y/, we computed F2' for all the vowels of the test, based on an algorithm presented elsewhere ([6]). Briefly, F2' is a non linear weighted sum of F2, F3, and F4. Similarly to our previous study on synthesized stimuli ([6]), it appears that F2' is a good normalizing parameter for the rounding contrast in French, for natural vowels uttered by speakers from 4 years old to adulthood. A category boundary of 15 Bark, revealed by a Probit modelling of the labelling function, allows the distinction between rounded vowels (F2' lower than 15 Bark) and unrounded ones (F2' higher than 15 Bark). This perceptual template is a good description of the speech task for the rounding contrast. However, our younger speaker, despite producing a focal /y/, can not achieve a perfectly intelligible /y/, that is, a rounded vowel with an F2' value lower than 15 Bark.

4. ARTICULATORY MODELLING

The case of the rounded vowel /y/ produced by the 4-year-old speaker is quite intriguing: while this vowel is produced with a convergence of F2 and F3, a value of F2' lower than 15 Bark is not achieved, resulting in a lower intelligibility. Simulations with an articulatory model are presented in order to study the acoustic effects of non uniform vocal tract growth on the /i/ versus /y/ contrast.

4.1. Overview of the model

For the present study, we have used the *Variable Linear Articulatory Model* (VLAM), developped by Shinji Maeda, a scaling of an adult version of Maeda's model ([7]) established from X-ray data and derived from a statistical analysis guided by knowledge of the physiology of the articulators. VLAM is controlled by seven parameters: protrusion and labial aperture; movement of the tongue body, dorsum, and tip; jaw and larynx height. A detailed description of VLAM can be found elsewhere ([6]).

4.2. Simulations

Simulations were performed on five growth stages, associated with representative vocal tract lengths from birth to adulthood. Since the method and the results are presented in [1], we shall focus, in the present study, on a description of the rounding contrast. For each growth stage, two sets of prototypes were generated and compared, based on two criteria: similar articulatory settings (articulatory prototypes) or similar acoustic relative position (acoustic prototypes). First, we situated the cardinal vowels /i/, /y/, /u/, and /a/ within the Maximal Vowel Space (MVS [7]), as defined by the dispersion-focalization theory ([8]): maximal F3 and F2, and minimal F1 for /i/, and close F2 and F3 (focal), and minimal F1 for /y/. The same acoustic prototypes were determined for the other growth stages.

Then, exploiting an inversion procedure, the articulatory commands, associated with the acoustic prototypes for the adult, were retrieved. These commands were generated in the younger vocal tracts, and are referred to as "articulatory prototypes". Figure 2 represents, in a newborn-like vocal tract, the articulatory prototypes (x), linked to the corresponding acoustic prototypes (stars) by a solid line. It appears that for /y/, the articulatory prototype does not correspond to a formant convergence. The difference between F2 and F3 is as high as 4 and 3.1 Barks, for the newborn and the 4-year-old speakers respectively. Contrastingly, the acoustic prototype for /y/ is associated with lower differences between F2 and F3 (around 1.6 Bark for the two speakers). It is worth noting that this low value conforms well with the F2-F3 distance calculated for natural vowels (see section 2.3). If we compare the simulations with the natural data, it appears that acoustic prototypes are produced by the speakers.

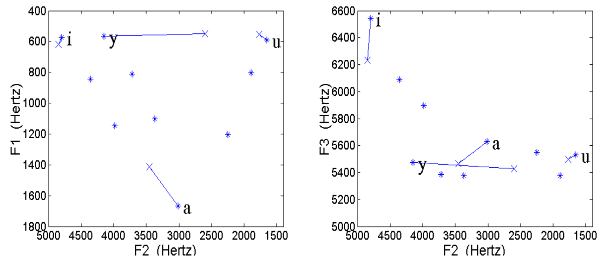

Figure 2: Representation of acoustic prototypes (labelled stars) and corresponding articulatory prototypes (represented by "x" and linked to the acoustic prototype by solid lines), for /i/, /y/, /u/, and /a/ simulated in a newborn-like vocal tract.

Focalization for /y/ is thus clearly not achieved when all articulatory parameters remain unchanged during growth. According to our previous study ([1]), realizing a focal /y/, for the young child, requires a fronting of the tongue body, compared to the adult male, in order to shorten the front cavity, and hence to increase its affiliated formant (F2). In turn, the back cavity is lengthened and F3 is lowered. As a consequence, F2 and F3 are closer to each other, realizing a focal vowel.

5. DISCUSSION

5.1. Focalization and adaptive articulatory strategy

Results presented so far showed that the focalization pattern for /y/, well realized by almost all the speakers, would require different articulatory strategies depending on vocal tract configuration. This strategy likely involves, for the youngest speaker, a more front position of the tongue body, compared to the adult male. One can suppose that the articulatory-to-acoustic links could therefore be different for this vowel. The strategy related to the tongue is preferred to a reduced lip protrusion gesture, in our simulation, since the latter articulatory gesture is a visible one, more easily reproduced by the child during the phase of imitation. Thus, lip protrusion is early associated with the phonological goal /y/. Figure 3 depicts the consequences, on the F2-F3 distance, of this compensation

ISBN 1-876346-48-5 © 2003 UAB

strategy. For the sake of clarity, formant trajectories are represented for the dynamic transition from /i/ to /y/, since the exchange of affiliation, in this case, becomes obvious.

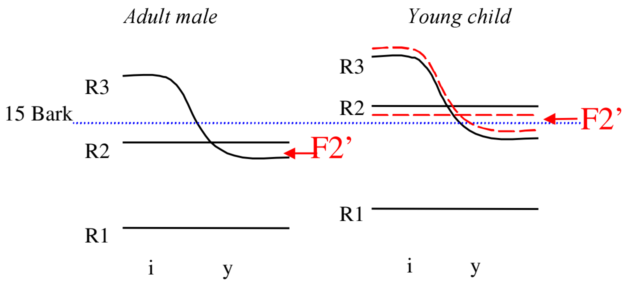

Figure 3: Schematic representation of the /iy/ transition, for the adult male and the young child, and of the effect of fronting the tongue body (solid line: acoustic effects of similar articulatory gesture; dashed line: acoustic effect of a fronting of the tongue body; dotted line: 15-Bark-category boundary of the invariant perceptual parameter for rounding in French (F2'); arrows: position of F2' according to the algorithm used in the present study).

5.2. Focalization at the detriment of intelligibility

For the young child, producing a focal /y/ can not be achieved without cost. Because of the overall shorter vocal tract, the pharynx and the oral cavity are shorter, and their resonant frequencies, higher. Figure 3 also represents the perceptual normalizing parameter F2' associated with the rounding contrast in French. The arrows stand for the F2' values for the two cases represented above, according to the F2' algorithm used in the current analyses. For the adult, reaching an F2' value lower than 15 Bark is possible together with a focalization of F2 and F3. The long cavities ensure low R2 and R3 values, and the F2 and F3 complex is sufficiently low to be below 15 Bark. On the contrary, the overall shorter cavities of the child result in higher R2 and R3. An F2' value lower than 15 Bark is achievable if no fronting of the tongue body (compensation strategy) is used. In this case, R2 is low, but R2 and R3 are too far apart to merge, and F2' corresponds to F2 (R3). However, realizing a focal vowel involves a close F2 and F3, and the articulatory strategy involved increases this whole formant group to a value above (or at the limit of) the 15-Bark-category boundary. Focalization plays such a crucial part in shaping the vowel system of speakers during growth that it is achieved at the cost of lower intelligibility for young children. With similar perceptual constraints (F2') and assuming the lip protrusion gesture to be defining the speech task as well for children, one can consider that vocal tract shapes representative of very young speakers prevent them from achieving perfectly intelligible /y/, in French. This formant convergence would allow the delimitation of the F2/F3 space, together with /i/ and /u/. Defining the limits of the F1 versus F2 as well as the F2 versus F3 space would be an acoustic objective guiding the speech task. These results shed light on the issue of production/perception relationships. Further investigations concerning the role of focalization in shaping the speaker's vowel system during growth are currently in progress.

5. CONCLUSIONS

The aim of the present paper was to describe the articulatory and acoustic variability observed in the realization of the rounding contrast in French, during non uniform vocal tract growth. Simulations with an articulatory model (VLAM) and recording of real data showed that focalization (close F2 and F3) is associated with the vowel /y/ produced by all but one of the speakers. For young children, an adaptation articulatory strategy is required in order to reach this goal, namely a fronting of the tongue body. However, this feature is realized at the cost of intelligibility.

ACKNOWLEDGMENTS

The author is indebted to Jean-Luc Schwartz and Louis-Jean Boë. Many thanks also to Marija Tabain.

REFERENCES

[1] L. Ménard, and L.-J. Boë (2000), "Exploring Vowel Production Strategies from Infant to Adult by Means of Articulatory Inversion of Formant Data", *ICSLP*, Beijing (China), 465-468.

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

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

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

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

[6] L. Ménard, J.-L. Schwartz, L.-J. Boë, S. Kandel and N. Vallée (2002), "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood ", *J. Acoust. Soc. Am. 111*: 1892-1905.

[7] Boë, L.-J. (1999), "Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults. Consequences for ontogenesis and phylogenesis", *ICSLP*, San Francisco, 2501-2504.

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