The development of lingual gestures in speech : an experimental approach to language development

Lucie Ménard* & Aude Noiray**

1. Introduction

Learning to speak a language is related to the emergence of sensorimotor "maps" in which vowels and consonants are associated with articulatory-acoustic vocal tract configurations. One major challenge for young children is to develop these associations while integrating anatomical changes, as well as motor, perceptual, and cognitive abilities (Green, Moore, & Reilly, 2002; Kuhl & Meltzoff, 1982; Vorperian et al., 2005). There is empirical evidence that the physical growth of the vocal tract is not complete until adolescence (Kent, 2004). Hence, from birth to adulthood, the production of vowels and consonants is likely to reflect continuous articulatory and acoustic adjustments, as the production system matures. Determining the exact role of each component (anatomical, motor, perceptual, and cognitive) for producing intelligible speech is a complex task, especially from an ontogenetic perspective.

The objective of this paper is to study the articulatory strategies used by children to produce speech targets. Those targets can be considered as phonological goals, implemented by phonetic articulatory gestures. Considering the facts that (i) 4-year-old children can produce intelligible phonemes, (ii) their motor control capacities are still immature, and (iii) their vocal tract anatomy greatly differs from that of adults, it is hypothesized that they use different phonetic strategies compared to adults to implement phonological targets. This report is part of a broader research program we have developed with our collaborators for the last decade (Ménard *et al.*, 2000; 2004). For the first time, this paper reports on articulatory data recorded via an ultrasound system. Data acquired by this method are compared to simulations with an articulatory-to-acoustic model (VLAM model), described below (Boë, 1999).

1.1. Non-Uniform Vocal Tract Growth, Motor Control Development, and Vowel Production

At the anatomical level, cineradiographic (Goldstein, 1980) and MRI data (Callan, Kent, Guenther, & Vorperian, 2000; Fitch & Giedd, 1999; Vorperian et

* Laboratoire de phonétique, Center for Research on Language, Mind, and Brain, Université du Québec à Montréal, Canada. Courriel : menard.lucie@uqam.ca
** Haskins Laboratories, USA.

al., 2005) have shown that the adult vocal apparatus is the result of a complex remodeling of the infant tract. At birth, the overall vocal tract, determined from the larynx to the lips, is approximately 8 cm long, whereas the adult male vocal tract is 17 cm long (Goldstein, 1980). According to Goldstein (1980), the pharynx/oral cavity ratio changes in terms of length from 0.5 at birth to 1.1 in adulthood for a male. Therefore, while the infants' pharyngeal cavity is much shorter than the oral cavity at birth, it is longer than the front cavity in adult mals (Fitch & Giedd, 1999). Various studies (Callan et al., 2000; Ménard, Schwartz, & Boë, 2004) have reported on the impact of these major anatomical modifications on speakers' productions of vowels at different developmental stages. For instance, Buhr (1980) suggested that the lack of [u]-like productions in infants' vowel inventories is related to the small size of their pharyngeal cavity.

In addition to the non-uniformity of infants' anatomical growth, the emergence of vocalic sounds is constrained by limited speech motor abilities in the first year of life (Davis & MacNeilage, 1995; Kent, 1976; Vilain, Abry, Badin, & Brosda, 1999). One main hypothesis is that consonant- and vowel-like sounds are produced with the tongue and the lips being passively raised and lowered via rhythmic motions of the jaw (MacNeilage & Davis, 1990). In this view, the vowels that do not occur in infants' early productions are those that cannot be articulated because of a lack of motor control over the tongue that would enable infants to produce tongue movements independent of jaw movements. The production of a wider variety of sounds occurs in late childhood or adolescence with the achievement of adult-like motor control (Kent, 1976; Smith & Zelaznik, 2004; Walsh & Smith, 2002). A few studies have investigated the development of articulatory control in young children. Among them, Green, Moore, Higashikawa, & Steeve (2000) showed that the coordination of lip and jaw movements develops as children grow up with a decrease in trial-to-trial variability indicative of the maturation of the speech production system. A similar pattern was observed by Smith and Zelaznik (2004) in a quantitative investigation of lip and jaw movements in 180 children and adults aged 4 to 22. Cheng, Murdoch, Goozee, and Scott (2007) have investigated tongue-jaw interactions with electromagnetic articulography in 48 subjects aged 6 to 38 during the production of the consonants /t/ and /k/. They observed an increase in synchronization of tongue-tip and jaw movements over time. Tongue-body and jaw movements did not show a similar pattern, but became less variable with age. Results were interpreted as indices of maturation of speech motor control with age. It seems that the acquisition of new controls and their integration into existing ones, such as the integration of tongue movements into already controlled jaw oscillatory movements, enables precision for the diversification of speech-related movements.

1.2. Previous Studies Combining Perceptual Tests and Articulatory Modeling

The effects of non-uniform cavity growth and motor control development on vowel production have been described in earlier studies of French (Ménard et al.,

2004) and American English (Ménard et al., 2009) using combined perceptual and modeling experiments. Five-formant vowels generated by an articulatory-to-acoustic model in vocal tracts representative of five growth stages (newborn, 4-year-old, 10-year-old, 16-year-old, and adult) were presented to French and Texan listeners in an identification task. Stimuli were generated on the assumption that all synthesized speakers exhibited adult-like motor control abilities, allowing the investigation of the effects of changes in vocal tract length alone. Even though the ratio of pharyngeal length to front cavity length differed across growth stages, acoustic targets for the three cardinal vowels /i u a/ could be perceived when synthesized in the infant's vocal tract. However, more stimuli were perceived as front and open vowels in the infant's vocal tract than in the adult male's. The vowels that are perceived to be front and open by French and English listeners correspond to those favored by the infant's vocal tract and to some of the most frequently produced vowels in infants' canonical babbling (Davis & MacNeilage, 1995). Further analyses allowed us to suggest language specific sensorimotor maps for French and English vowels from birth to adulthood. The analysis of these maps suggested that different ranges of articulatory parameters were used, depending on vocal tract morphology. For instance, a given perceived French or English vowel might be produced with different positions of the jaw, tongue, and lips in the first months of life compared to the adult stage. Those conclusions, however, were exclusively based on modeling studies and did not include natural vowels. In the present paper, we explore the development of tongue gestures using articulatory data recorded by means of an ultrasound system.

1.3. Using ultrasound to image tongue movement

Ultrasound imaging has become increasingly popular in speech research (Adler-Block, Bernhardt, Gick, & Bacsfalvi, 2007; Bernhardt, Gick, Bacsfalvi, & Adler-Bock, 2005; Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Stone, Epstein, & Iskarous, 2004; Whalen, Iskarous, Tiede, Ostry, Lehnert-LeHouillier, & Vatikiotis-Bateson, 2005). One of the main advantages of this technique is that it provides a global view of the tongue contour in the mid-sagittal or coronal plane, whereas flesh point tracking methods such as EMMA (electromagnetic mid-sagitttal articulography, Perkell et al., 1992) or x-ray microbeam (Westbury, 1994) provide the coordinates of the receivers attached to the tongue. Figure 1 provides an illustration of an experimental setup using ultrasound imaging (Figure 1.a.) with the mid-sagittal edge of the tongue extracted (Figure 1.b.). The tongue edge is graphically represented with the tongue tip on the right and the root on the left as in all subsequent figures. Since high-frequency sound waves are not reflected by bony structures, jaw shadow and hyoid shadow sometimes appear on the image, as shown in Figure 1b. Ultrasound is not invasive unlike EMMA which requires the presence of a receiver(s) in the subject's mouth. Instead, tongue movement data are collected via a transducer positioned under the subject's chin. Also, the pre-recording and familiarization phases are shorter during ultrasound experiments, since the preliminary gluing of receivers on subjects' tongues and calibration trials are not required. As a result, ultrasound is a non-invasive, affordable and possibly portable technique that makes it very suitable for studies with clinical populations and young children for whom data collection cannot be as controlled as it is in laboratory settings.

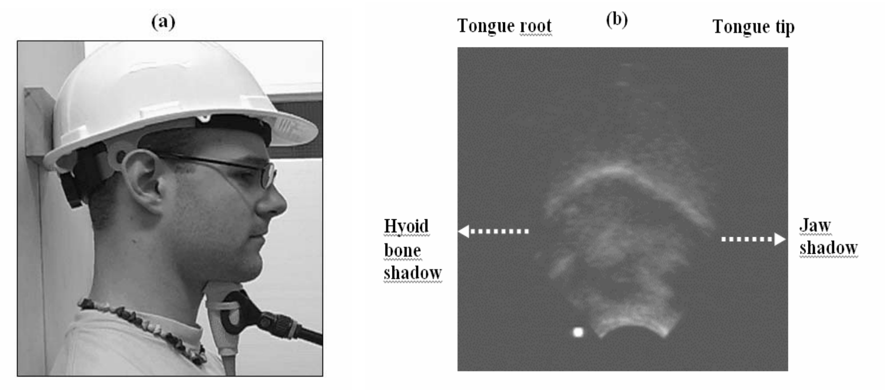


Figure 1: Experimental setup (a) and mid-sagittal tongue contour (b) collected with ultrasound imaging. The tongue tip is at the right and the tongue root is at the left. The arrows represent the jaw shadow and the hyoid bone shadow.

2. METHOD

In this study, we used a method in which ultrasound mid-sagittal images of the tongue were compared with respect to their shapes and positions (Ménard et al., 2008; Aubin & Ménard, 2006). To this aim, tongue contours were reshaped into a triangle. Various measures of angles and distances were computed. The method was applied to tongue contours simulated by an articulatory-to-acoustic model as well as natural vowels.

2.1. Simulations using the VLAM articulatory model

The *Variable Linear Articulatory Model* (VLAM), developed by Shinji Maeda, is a scaled version of Maeda's model for adults (Maeda, 1979) designed from cineradiographic data and derived from a statistical analysis guided by knowledge of the physiology of the articulators. VLAM has been extensively described in Boë (1999) and 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). Each articulatory parameter can be adjusted to a value ranging from -3.5 to 3.5 (Those values correspond to the standard deviations around the mean calculated from the x-ray images). From the values of these seven parameters, a two-dimensional mid-sagittal section is generated, 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. VLAM has been compared to real data (Ménard et al., 2004), and it actually generates realistic articulatory and acoustic vowel configurations. Overall vocal tract lengths and cavity lengths are in line with MRI measurements (Fitch & 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 participants aged 3 to adulthood (Lee et al., 1999; Hillenbrand et al., 1995). This procedure is thus well suited to modeling vowel production.

For the purpose of the present study, the model was set to two growth stages: 4 years old and adulthood (21 years old). The age 21 corresponds to the mature stage in the model, when the growth process is complete (Goldstein, 1980). The overall vocal tract length obtained for the 4-year-old and the adult vocal tract were 10.67 cm and 17.45 cm respectively, and the ratios of back and front cavity length were 0.87 and 1.1 respectively. An example of a mid-sagittal contour, from the larynx to the lips, for a neutral configuration, is depicted in Figure 2. For the sake of clarity, the contour is superimposed on an image of the skull. Mid-sagittal contours corresponding to the three French vowels /i/, /u/, and /a/ were generated. Optimal formant triplets (F_1, F_2, and F_3) were first determined, based on acoustic criteria inspired from the dispersion-focalization theory (DFT, Lindblom, 1996; Schwartz et al., 1997). The validity of these formant targets has been confirmed in previous studies (Ménard et al., 2002). Because of the many-to-one relations between articulatory configurations and acoustic targets, many articulatory configurations, and thus, many mid-sagittal contours, can be associated with each of these formant triplets. The method employed is an inversion procedure consisting of the pseudo-inverse of the Jacobian matrix (Jordan & Rumelhart, 1992). We ensured that the chosen articulatory configuration for each vowel corresponded to earlier descriptive articulatory and acoustic studies of French vowels (Vallée, 1994; Bailly et al., 1995). From 55 to 65 mid-sagittal contours were generated for each of the three French vowels.

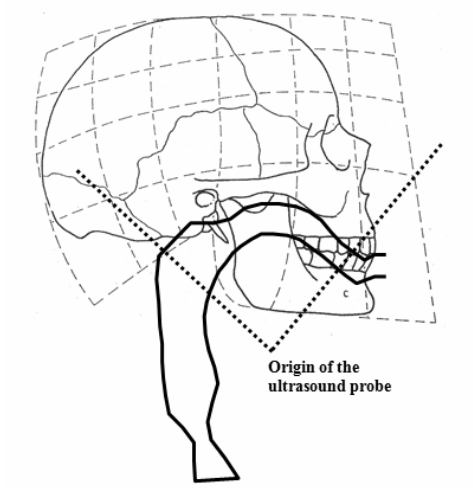


Figure 2: Synthesized mid-sagittal contour in neutral position (thick solid line) superimposed on a skull. The thick dotted line corresponds to the edges of the ultrasound transducer scanning area.

2.2. Analysis of natural vowels: experimental setup

One 4 year-old female child and one 24-year-old adult were recruited for a production experiment. The 4-year-old speaker is referred to as CF and the adult speaker as AF. Both participants were native speakers of Canadian French. A screening procedure ensured that no subject would show any hearing or phonological disorders. Stimuli consisted of 10 repetitions of the syllable /V_1CV_2/ in which V corresponded to one of the three cardinal vowels /i u a/ and C to one of the three consonants /p t k/. In the present study, analyses were performed on V_1 only. Subjects were recorded with a combined ultrasound (Sonosite 180Plus system with a 84-degree probe) and audio system, in a sound booth at the *Laboratoire de phonétique* at UQAM, in Montreal. They were seated comfortably in a chair and the ultrasound probe was held under their chin by the experimenter. While head stabilizing devices are usually used to hold the head and the probe stable during the experiment, it is not possible to use similar methods with young subjects. Therefore, the same experimental setup was used for both subjects to avoid methodological bias. The acoustic signal was collected with a unidirectional microphone. Both ultrasound and speech sound signals were recorded on a miniDV Panasonic AG-DVC 30 camcorder, in NTSC format. The sampling frequencies were 29.94 fps for the video signal and 22050 Hz for the audio signal.

2.3. Data analysis

For each of the mid-sagittal synthesized views (section 2.1), a section of the tongue contour was extracted in order to simulate an ultrasound scan, as shown in Figure 2. To ensure an appropriate selection (in length and in space), an 84-degree angle was superimposed on the mid-sagittal view, corresponding to the visualization angle of the Sonosite 180Plus ultrasound transducer. Guided by x-ray images, the origin of the angle (representing the origin of the transducer) under the chin, close to the neck was located, as shown in Figure 1a. The thick dotted lines in Figure 2 illustrate the edges of the selected area of the mid-sagittal section. The selected tongue contour was sampled at 100 points. The superimposed tongue contours for each vowel are presented in Figure 3.

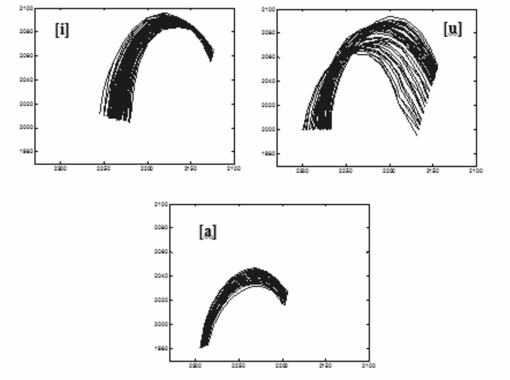


Figure 3: Mid-sagittal contours generated by the VLAM articulatory model, for the six French vowels [i u a] generated in an adult male vocal tract.

Regarding natural vowels, ultrasound images corresponding to the vowels' acoustic midpoint were extracted with Adobe Premiere Pro. Tongue surface contours were measured using EdgeTrak, a semi-automatic system designed for the extraction and the tracking of tongue contours (Li et al., 2003). Tongue contours were then sampled at 100 points.

All tongue contours (synthesized and natural) were exported to a homemade Matlab application, Lingua, which allows for the extraction of several parameters quantifying tongue contours. In a first step, a 19-segment radial grid with an inter-segment angular distance of 5 degrees was superimposed on each contour. This point was considered to be the origin of the transducer. For each contour, the intersection point between tongue contour and each segment of the grid was calculated. Radial distances between this point and the origin of the grid were calculated for each contour, allowing a description of tongue motion along various segments and across tokens. The x and y coordinates of the highest point (on the vertical axis) of the contour were also extracted. Note that this point can be interpreted as an absolute measure of tongue height, but does not correspond to a measure of tongue height relative to the palate.

In a second step, each contour was reshaped into a triangle. Figure 4 shows examples of the parameters analyzed in the present study. For the sake of clarity, the tongue contour was delineated by a solid line and the palate trace by a thin dotted line. The triangle base was determined via the first and last intersection points between the contour and the grid lines (thick dashed line AB on the triangle). The peak of the triangle (point C on Figure 4) represents the highest point of the tongue contour with respect to the triangle base. Angular measures of CAB, CBA, and ACB were extracted, and measures of tongue curvature and tongue curvature position were determined from points A to D as represented in Figure 4. Tongue curvature was defined as the ratio of the distance CD over the distance AB. As seen in Figure 4, contour (b) had a larger value for tongue curvature than contour (a). Tongue curvature position was defined as the ratio of the distance AD over the distance DB. Contour (b) in Figure 4 had a smaller value for tongue curvature position than contour (a). Curvature position is a measure of the position of the mass of the tongue relative to the whole tongue, whereas the x coordinate of the highest point of the tongue can be considered as an absolute measure of front-back dimension. Tongue curvature degree and tongue curvature position have proven to be robust in probing movement (Ménard et al., submitted) and can thus be used to characterize tongue shapes collected in such an experimental setup.

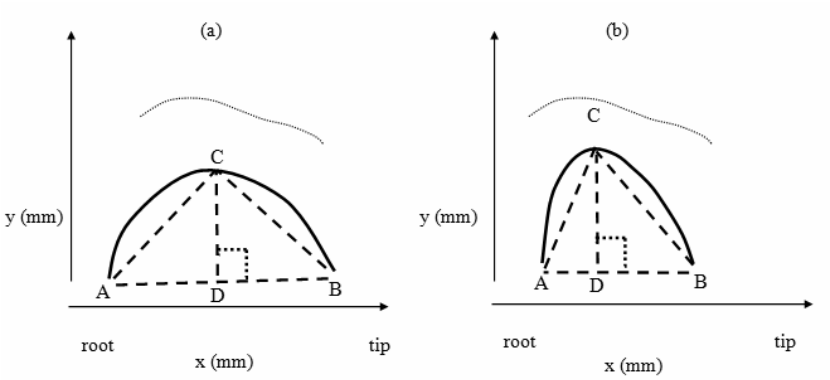


Figure 4: Schematic representations of two mid-sagittal contours in (x,y) coordinates and parameters extracted by the analysis in Lingua.

3. RESULTS AND DISCUSSION

The various parameters extracted by Lingua were compared across vowels, for both the synthesized contours and the natural contours. Synthesized contours were plotted in multidimensional spaces to evaluate the extent to which anatomical differences between 4-year-olds and adults influence the partition of the tongue curvature * tongue position space. Note that only articulatory parameters related to tongue position were analyzed in this study. Several distinctions that were not captured by ultrasound imaging can indeed be implemented in the labial dimension. Also, tongue shape and position were described for natural vowels and compared to the synthesized ones.

3.1. Synthesized vowels

Values for tongue curvature degree and tongue curvature position for the three French vowels are presented in Figure 5, for the 4-year-old vocal tract (left panel) and the adult vocal tract (right panel). Figure 5 shows that the three cardinal vowels [i u a] are clearly differentiated within this two-dimensional space for both growth stages. Along the dimension of tongue curvature position, [i] corresponds to a higher value than [u], which in turn corresponds to a higher value than [a]. This axis is thus closely related to the front-backness of the tongue, a lower value being related to a more back tongue position. As for tongue curvature degree, vowels like [i] and [a] contrast with [u], for instance, the latter corresponding to higher values of tongue curvature degree than the former. This dimension reflects the degree of flatness or bunching of the tongue, more bunched tongue contours having higher values of tongue curvature degree. The pattern observed in this study corroborates the results of previous articulatory investigations (Savariaux et al., 1995) using x-ray technology, which demonstrated this very particular tongue shape typical of [u]. This shape is indeed necessary to create two Helmholtz resonators within the vocal tract, affiliated with low values of F1 and F2. Thus, tongue curvature degree and tongue position allow for a reliable characterization of tongue bunching/flatness and tongue front/backness.

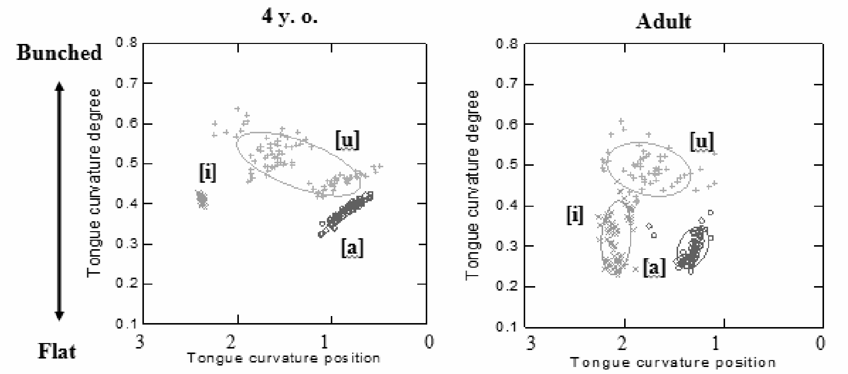


Figure 5: Values of tongue curvature degree and tongue curvature position for synthesized tongue shapes generated by VLAM for the three cardinal vowels [i u a]. Right panel: 4-year-old vocal tract; left panel: adult vocal tract.

Developmental differences can also be observed in Figure 5. While the ellipsis for the vowel /i/ is rather small for the 4-year-old vocal tract, the corresponding ellipsis is quite large for the adult vocal tract. This pattern is in line with our previous claims (Ménard *et al.*, 2004) that the morphology of the young speaker's vocal tract constrains the production of the vowel /i/ in the tongue shape dimension. On the contrary, the range of values for the tongue curvature degree corresponding to the vowel /u/ in the 4-year-old vocal tract is larger than it is for the adult vocal tract. This large /u/ dispersion leads to a reliable differentiation along the tongue curvature degree dimension between /u/ and /i/ for the adult vocal tract, whereas those vowels are not well differentiated in the 4-year-old vocal tract.

The developmental differences discussed above are those predicted by a model simulating the sole effect of vocal tract length differences on these sets of measurements. Such simulations are extremely precious as they can be used to interpret the patterns observed in natural vowels and disentangle the roles played by anatomy and motor control differences in such vowels.

3.2. Natural vowels

Dispersion ellipses of the three vowels in the tongue curvature degree vs. tongue curvature position space for the 4-year-old speaker CF are plotted in Figure 6. Data for the adult speaker are plotted in Figure 7. Vowels uttered in [VpV] context are presented on the upper left panel, vowels uttered in [VtV] context are plotted on the upper right panel and vowels produced in [VkV] context are presented on the lower panel. Several conclusions can be drawn from these results. First, there was a significant effect of the consonantal context on tongue shape parameters. Indeed, the location and dispersion of the ellipses for a given vowel varied a lot from one context to another for both speakers[1]. An important difference between speakers CF (Figure 6) and AF (Figure 7) could be observed in the size of the dispersion ellipsis associated with the three vowels. While the ellipsis associated with [u] was larger than [i] and [a] for the adult AF, the corresponding ellipsis was rather similar in size to [i] and [a] for the child speaker CF (less so for the [VkV] context). It is unlikely that such a pattern is due to differences in vocal tract morphology since Figure 5 suggests that, with similar motor control skills, the 4-year-old would produce a larger dispersion ellipsis for [u] compared to the adult speaker. Indeed, the very small pharynx compared to the oral cavity, typical of the 4-year-old vocal tract synthesized in Figure 5, allows more variability in the size of the dispersion ellipsis associated to [u] in the tongue shape dimensions. The reverse pattern is however observed for natural vowels and rather suggests that motor control refinement allowed the adult speaker to exploit wider ranges of tongue curvature degree and position for [u] compared to the child speaker.

---

[1] Coarticulation effects are currently investigated and will be discussed in another paper.

Figure 6: Values of tongue curvature degree and tongue curvature position for tongue shapes obtained for the 4-year-old female speaker CF. Higher left panel: initial vowels in [VpV] context; higher right panel: initial vowels in [VtV] context; lower panel: initial vowels in [VkV] context.

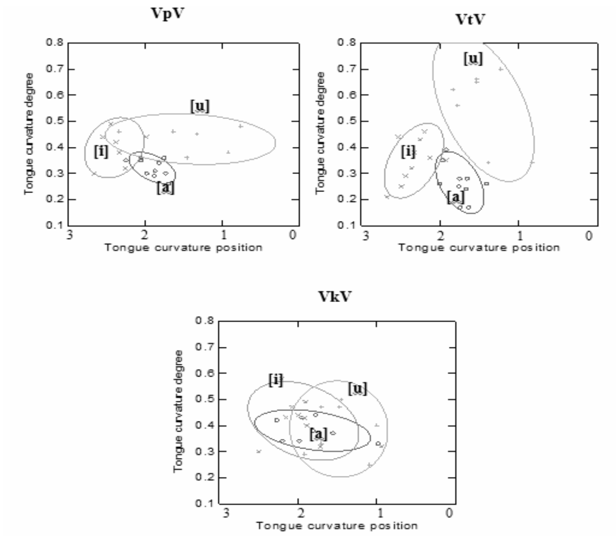
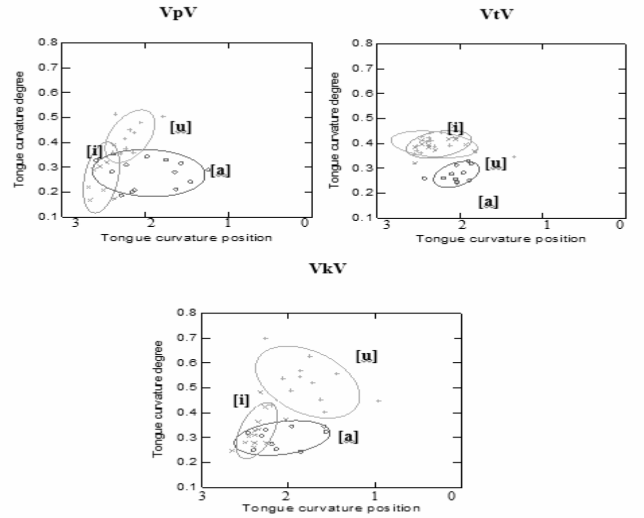


Figure 7: Values of tongue curvature degree and tongue curvature position for tongue shapes obtained for the adult female speaker AF. Higher left panel: initial vowels in [VpV] context; higher right panel: initial vowels in [VtV] context; lower panel: initial vowels in [VkV] context.



Another difference between the adult AF and the child CF speakers can be seen in the amount of overlap between the dispersion ellipses corresponding to the three vowels in [VtV] and [VkV] contexts. While the vowels [i] and [u] produced by CF (Figure 6, upper right graph) show great overlap, those vowels are well differentiated in speaker AF (Figure 7). The reverse pattern can be observed between CF and AF between vowels [i], [u], and [a] in [VkV] context. This overlap in articulatory dimensions between different vowel categories can reflect coarticulation, especially in the context of alveolar and palatal consonants, for which tongue movements is required to produce both the vowel and the consonant. These results are in line with Sussman et al. (1999) who showed that in a CV sequence in which C is an alveolar, the developmental course consists in gradually differentiating between the alveolar and dorsal tongue gestures, resulting in reduced coarticulation over time. On the contrary, in CV contexts in which C is a palatal, children must learn to integrate two sounds produced with the same articulator, presumably leading to increased coarticulation.

To summarize, the combination of articulatory-to-acoustic simulations and articulatory recordings provide a unique database on the phonetic implementation of phonological targets. We are currently collecting more data on larger groups of subjects. Apart from their contribution to theories of language acquisition and development, such studies provide normative data in speech production and can be used as a baseline for clinical experiments.

4. CONCLUSION

In this paper, we conducted an exploratory study of the development of lingual gestures in French. Both datasets, one with an articulatory model and the other with natural data collected with an ultrasound system, were analyzed. In line with our previous studies (Ménard et al., 2002; 2004), it was shown that a 4-year-old speaker could produce well differentiated tongue shapes in the tongue curvature position vs. tongue curvature degree for the three cardinal vowels /i u a/. On this account, a relatively small pharyngeal cavity does not prevent the speaker from producing contrasted tongue positions.

The analysis of natural vowels produced by a 4-year-old and an adult female using an ultrasound system has provided evidence that differentiation of [i u a] in the tongue curvature degree and position space depends on consonantal context and growth stage. Development trends are thus constrained by anatomy.

Of course, the data presented in this study comes from a limited dataset (two speakers only); that will be increased with more data collected from additional speakers. We are currently working on a larger corpus to investigate the development of coarticulatory patterns in French. Nevertheless, the present study confirms the validity of ultrasound to study tongue gestures in young subjects. Furthermore, the use of the tongue curvature degree and tongue curvature position parameters to characterize tongue contours proves well-suited to studying tongue edges. We believe the combined use of articulatory modelling and ultrasound imaging can shed light on speech development.

REFERENCES

Adler-Block M., Bernhardt B. M., Gick B. & Bacsfalvi P., 2007, The Use of Ultrasound in Remediation of North American English /r/ in 2 Adolescents. *American Journal of Speech-Language Pathology* 16, p. 128-139.

Bernhardt B., Gick B., Bacsfalvi P. & Adler-Bock M., 2005, Ultrasound in Speech Therapy with Adolescents and Adults, *Clinical Linguistics and Phonetics* 19, p. 605-617.

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 international congress of phonetic sciences 3*, San Francisco, p. 2501-2504.

Bock A., Bernhardt B., Gick B., & Bacsfalvi P., 2007, The use of Ultrasound in remediation of North American English /r/ in 2 Adolescents, *American Journal of Speech-Language Pathology* 16, p. 128-139.

Buhr R. D., 1980, The emergence of vowels in an infant, *Journal of Speech, Language and Hearing Research* 23, p. 73-94.

Davis B. L., & MacNeilage P. F., 1990, Acquisition of correct vowel production: a quantitative case study, *Journal of Speech, Language and Hearing Research* 33, p. 16-27.

Davis B. L., & MacNeilage P. F., 1995, The articulatory basis of babbling, *Journal of Speech, Language and Hearing Research* 38, p. 1199-1211.

Fitch W. T., & Giedd J., 1999, Morphology and development of the human vocal tract, *A study using magnetic resonance imaging* 106, p. 1511-1522.

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

Green J. R., Moore C. A., & Reilly K. J., 2003, Methodological issues in studies of early articulatory development : A response to Dworkin, Meleca, and Stachler, *Journal of Speech*, Language, and Hearing Research 46, p. 1020-1021.

Kent R. D., 1976, Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies, *Journal of Speech and Hearing Research* 19, p. 421-447.

Kuhl P. K, & Meltzoff A., 1982, The bimodal perception of speech in infancy, *Science 218*, p. 1138-1141.

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

Li Min, Kambhamettu C, & Stone M., 2003, EdgeTrak, a program for band-edge extraction and its applications, *Sixth IASTED International Conference o Computers, Graphics and Imaging*, Honolulu, HI.

MacNeilage P. F. & Davis B., 1990, Acquisition of speech production: Frame, then, content, in M. Jeannerod (ed), *Attention and performance XIII*, Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, p. 453-475.

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, p. 465-468.

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, p. 1059-1080.

Stone M. L., Epstein M. A. & Iskarous K., 2004, Functional segments in tongue movement, *Clinical Linguistics and Phonetics* 18, p. 507-521.

Sussman H. M., Duder C., Dalston E., & Cacciatore A., 1999, An acoustic analysis of the development of CV coarticulation : A case study, *Journal of Speech and Hearing Research* 42, p. 1080-1096.

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 childhood: A magnetic resonance imaging study, *The Journal of the Acoustical Society of America* 117, p. 338-350.

Whalen D. H., Iskarous K., Tiede M. K., Ostry D. J., Lehnert-LeHouillier H., Vatikiotis-Bateson E. et al., 2005, HOCUS, the Haskins Optically-Corrected Ultrasound System, *Journal of Speech, Language, and Hearing Research* 48, p. 543-553.