



Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses

Guillaume Barbier¹, Pascal Perrier¹, Lucie Ménard², Yohan Payan³, Mark K. Tiede⁴,
Joseph S. Perkell^{5,6}

¹Speech and Cognition Department, GIPSA-lab & Grenoble University, Grenoble, France

²Department of Linguistics, Université du Québec à Montréal, Montréal, Québec, Canada

³Univ. Grenoble Alpes, TIMC-IMAG, F-38000 Grenoble, France
CNRS, TIMC-IMAG, F-38000 Grenoble, France

⁴Haskins Laboratories, New Haven, Connecticut, USA

⁵MIT, Cambridge, Massachusetts, USA;

⁶Boston University, Boston, Massachusetts, USA

guillaume.barbier@gipsa-lab.grenoble-inp.fr

Abstract

This study investigates speech motor control in 4-year-old Canadian French children in comparison with adults. It focuses on measures of token-to-token variability in the production of isolated vowels and on anticipatory extra-syllabic coarticulation within V_1 -C- V_2 sequences. Acoustic and ultrasound articulatory data were recorded. Acoustic data from 20 children and 10 adults have been analyzed. Thus far, ultrasound data have been analyzed from a subset of these participants: 6 children and 2 adults. In agreement with former studies, token-to-token variability was greater in children than in adults. Strong anticipation of V_2 in V_1 was found in all adults, but not in children. Most of the children showed no anticipation at all and some of them showed a small amount of anticipation along the antero-posterior dimension only, manifested in the acoustic F2 dimension. These results are interpreted as evidence for the immaturity of children's speech motor control from two perspectives: insufficiently stable motor control patterns for vowel production, and a lack of effectiveness in anticipating forthcoming gestures. In line with theories of optimal motor control, anticipatory coarticulation is assumed to be based on the use of internal models of the speech apparatus and the increasing maturation of these representations as speech develops.

Index Terms: speech production development, speech motor control, coarticulation, planning.

1. Introduction

This project aims to use articulatory and acoustic observations of children's speech to characterize the maturity of the neural representation of speech sounds and the speech motor system. This paper extends our previous study published at Interspeech 2013 [1]. It includes more participants and offers a more comprehensive statistical analysis of the experimental data. In this introduction the rationales of our approach, in particular of the choice of indices to assess speech motor control maturity, are explained.

The findings relative to the development of the motor control of the limbs [2-5] and of the lips and the jaw [6-8] are quite clear: the acquisition of fine motor control is a protracted process that seems to be fully accomplished only in late adolescence [9]. According to the literature, the measure of the variability in repetitions of a single task is a quite robust index

of motor control stability. However, few studies have used this index to assess tongue motor control in children [10-11]. Additional evaluation is required, and one purpose of this study is to contribute to such validation and to extend this measure to very young children.

Another characteristic of skilled motor control is the capacity to plan and organize movements over specific sequences of goals. In speech, this function can be assessed via the amount of anticipatory coarticulation. Unfortunately, contradictory results have been found in past studies of the development of lingual coarticulation [10-16]. Some possible reasons for these contradictory results could be the often small number of participants, the use of acoustic measurements only (except a few recent studies using ultrasound (US) [10,11,16]), the differences in the selected acoustic variables, the differences in the speech corpora under investigation, and the large spread in age groups. To contribute to this debate for young children, we designed a study combining articulatory (2D midsagittal US tongue imaging) and acoustic data, focused on a narrow age group (from 4 years to 4 years 11 months) and involving a substantial number of participants (20 children, 10 adults).

In our view, focusing on a narrow age group is crucial, because of the evolution of the status of the syllable across ages in early speech development. According to the theoretical framework proposed in particular by MacNeilage or Nittrouer [17,13], syllables (or even larger units, Vihman [18]) could be considered as the basic speech units in infant's speech. Phonemes would then gradually emerge from these more global patterns to become over time the main speech units in young children. Accordingly, we can first expect to observe in speech development a strong intra-syllabic cohesion associated with strong coarticulation in infants, and then a gradual decrease of this cohesion as phonemes emerge. At a final stage, an increase in intra and extra-syllabic coarticulation associated with serial-order motor control planning based on phonemes should be observed when children are mastering the neural representations of phonemes and the spatio-temporal organization of speech gestures.

In this context, this study investigates the emergence of extra-syllabic anticipatory coarticulation. Indeed, it has been suggested that anticipatory coarticulation would mainly be the consequence of gesture planning [19]. In line with classical hypotheses on the planning of serial-order motor tasks [20] we assume that motor speech planning involves the capacity to predict the effect of motor commands on articulatory

movements and on sound, and the ability to integrate these predictions to find the most adapted motor commands for the correct achievement of the speech goals. The focus of our study is 4-year-old children, mainly for two reasons: (1) numerous publications have suggested that at this age, phonemic representations exist [21-23], allowing us to infer that speech goals are phoneme related; (2) motor control studies have shown that at this age a key initial step is taken in the acquisition of the neural representations of the motor system, which enables predictions of the link between motor commands and goals and opens the door to the emergence of adult-like motor control strategies [24-27]. Beyond this, the current study bypasses complex debates about the nature of the basic speech units during development.

In sum, anticipatory coarticulation is assumed to be based on the use of internal representations of the speech apparatus and we consider its efficiency to reflect the maturity of these representations. Presumably, mature speakers have acquired implicit knowledge of the amount of produced variability that is compatible with correct perception of the produced sounds by listeners. Mature speakers use this tolerance of variability to plan and execute a sequence of speech gestures with minimized articulatory effort [28]. Our hypothesis is that 4-year-old children do not have sufficient experience with the sensory consequences of motor acts to be able to implement this effort-minimizing strategy effectively, particularly with respect to variability in sound categories. As a consequence, we predict that children's abilities to plan upcoming gestures could be either limited or inaccurate. This hypothesis is tested here through the analysis of extra-syllabic anticipatory coarticulation by comparing the performances of 4-year-old children to those of adults.

2. Material and Methods

2.1. Participants

Twenty 4-year-old Canadian French children (4 years 0 months to 4 years 11 months) and 10 Canadian French adults (19-30 years old) were recruited in Montréal. Canadian French was the first language of all participants. All children lived in monolingual French families and were educated in French only. Participants reported no history of speech or hearing problems. All participants showed normal audition, by passing a bilateral pure tone screening test at 20dB at 250Hz, 500Hz, 1000Hz, 2000Hz and 4000Hz before the experiment. All participants and participant's parents, in the case of children, were informed about the procedures before the experiment and gave their consent. This study was approved by the ethical committee of the Université du Québec à Montréal (UQÀM). This paper presents the acoustic results for all participants and the articulatory results for 6 children and 2 adults.

2.2. Data acquisition

Ultrasound is a noninvasive imaging technique. It is suitable for use with very young children, and provides a real-time 2D view of most of the tongue, with good temporal (15Hz-200Hz) and spatial (~1mm) resolution. To obtain reliable measurements of tongue movements corrected for any probe or head movements, we used the HOCUS system (Haskins Optically Corrected Ultrasound System, [29]). HOCUS uses optical tracking (Optotrak, NDI Certus) of infrared emitting

diodes (iREDs), positioned both on the US probe and on the head of the participant, to provide a representation of the data in a movement-corrected head-centric frame of reference. This approach is appropriate for developmental studies, in that it preserves some freedom of movement for the participants. In this study, an iRED was also placed on the chin to allow tongue movements to be dissociated from jaw movements by providing an index of jaw motion.

Synchronous recordings of tongue movements in the midsagittal plane (at NTSC 29.97 Hz) and of the speech signal (at 44.1kHz) were made by the US device (Sonosite 180 Plus) and a directional microphone. The Optotrak system was used to record audio and the positions of the iREDs concurrently. The two data types were synchronized in post-processing by cross-correlating the two audio signals. After head-movement correction and alignment to a coordinate system centered on the upper incisors, data were mapped onto a 3D view in which iREDs' positions and the tongue imaging plane were visible.

2.3. Task

Data were collected on-site at Montréal day care centers and at the Laboratoire de Phonétique, UQÀM. Participants were seated in front of the Optotrak, which was disguised as a puppet theater, and the US probe was held under their chins by a microphone stand. One experimenter checked that head of the speaker was not moving much with reference to the US probe, and that most of the tongue was visible; another experimenter controlled the recording (Optotrak and US) and checked that all the iREDs were visible during the trials.

The corpus consisted of 8 to 10 repetitions of isolated vowels /i e ε a u/ and of V_1 -C- V_2 sequences with C as /b d g/, V_1 as /ε a/, for which a certain variability was expected, and V_2 as /i a/, which correspond to two extreme front/back and high/low articulations. Isolated vowels were used to measure the dispersion of vowel production in the F1-F2 plane associated with token-to-token variability. The V_1 -C- V_2 sequences were designed to measure the effects of the anticipation of V_2 in V_1 .

The task was presented as a puppet game, with a third experimenter serving as puppet master. The puppets' names were isolated vowels or V_1 -C- V_2 sequences. Puppets were presented in different pairs. The order of appearance was randomized. The task was to pronounce the name of the puppet when it appeared. Thus, participants had to recall, plan and execute a speech gesture or a sequence of speech gestures.

2.4. Data processing and statistical analyses

2.4.1. Acoustics

The acoustic signal was downsampled to 16 000 Hz in order to achieve more accurate formant detection. This signal was first labeled manually with *Praat* [30]: for vowels V_1 and V_2 , the beginning of the vowel was defined as the first descending zero-crossing of the signal after the clear emergence of F2, and the end of the vowel was defined as the first descending zero-crossing after the disappearance of F2. Automatic formant detection in the midpoint of the vowels was carried out with a Linear Predictive Coding (LPC) method. Because formant tracking is difficult in child speech, with the potential risk for detection errors, we combined the measure of the frequencies of the maxima in the LPC spectra with the measure of the frequencies of the poles in the LPC filter. For each vowel, a

range of acceptable formant values was used to guide the selection of the right formants among all possible candidates, and to remove outliers.

Prior to the statistical analysis, F1 and F2 values were converted to z-scores for each speaker separately, in order to reduce inter-speaker variability. This transformation (acting like a vowel-space normalization across speakers and across ages) enabled us to group children’s z-scored formant values and adult’s z-scored formant values and to compare adults and children on this basis.

2.4.2. Ultrasound data

As concerns articulatory data, the US images corresponding to the midpoint of the vowels were used. The midsagittal tongue contour was extracted using a semi-automatic procedure, *GetContours* (Haskins labs), similar to other edge extraction tools such as *EdgeTrak* [31]. Contours were converted to 3D head-centric coordinates using the HOCUS procedures described above. Ultrasound data were converted to polar coordinates in order to perform statistical tests. The use of polar coordinates facilitated the comparison of the tongue contours across conditions.

2.4.3. Statistical analyses

The statistical analyses of US data were based on the SS ANOVA method [32], widely used in speech production studies (e.g. [33]). This particular method of functional analysis consists of approximating the data by spline functions, and supports comparison of sets of data by constructing confidence intervals. In this study, we used 95% confidence intervals, corresponding to a $p = 0.05$ threshold. Since there is, to our knowledge, no commonly accepted method for comparing tongue shapes among speakers, we only compare tongue curves subject-by-subject.

The statistical analysis of the acoustical data includes a descriptive analysis of the F1 and F2 values of the isolated vowels, split by age groups (adults *versus* children). For the V_1 -C- V_2 sequences, in order to examine the influence of vowel V_2 on the F1 and F2 values of vowel V_1 , we conducted statistical analyses based on a General Linear Mixed-Effects Model [34] under the R environment [35]. Since this implementation does not support exact computations of *p-values*, we use a function (*pamer.fnc*, [36]) that provides upper and lower estimates of *p-values*. A threshold of $p = 0.01$ has been used in this study. Pair-wise post-hoc comparisons have been done with the *glht* function [37].

3. Results

3.1. Token-to-token variability

The distributions of the z-cored F1 and F2 values of isolated vowels /i e ε a u/ are displayed in Fig. 1 and 2, for adults and children respectively. Token-to-token variability is measured by the standard deviation within each age group. We consider this variability to reflect the stability of speech motor control for vowel production: The smaller the standard deviation, the greater the stability of the control. These figures show a clear trend for children to present more variability than adults. More quantitatively, the mean standard deviation, across all vowels and all speakers, is 2.13 times greater in children than in

adults. Thus, we conclude that the stability of the control is greater in adults than in children.

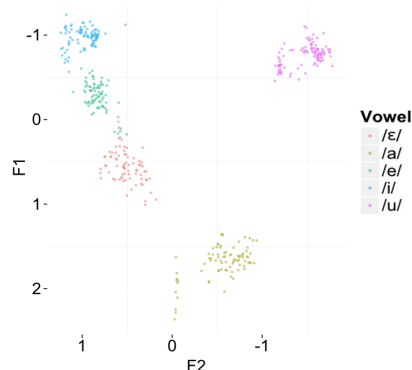


Figure 1: Token-to-token variability of isolated vowels for all adults in the z-scored (F2, F1) plane.

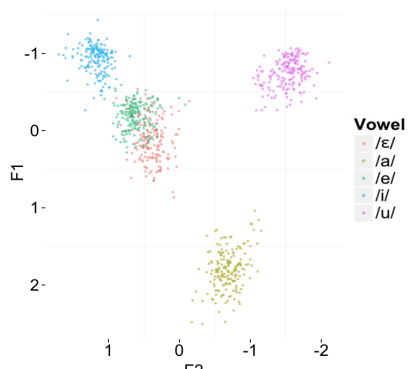


Figure 2: Token-to-token variability of isolated vowels for all children in the z-scored (F2, F1) plane.

3.2. V_1 -C- V_2 sequences

Figures 3 and 4 present, for adults and children respectively, the distributions of the z-scored F1 and F2 values of vowels V_1 (a/ and /ε/) within the V_1 C V_2 sequences for two different vowels V_2 , namely /i/ and /a/. In these figures, the effect of V_2 on V_1 can be measured through the difference in V_1 formant values depending on the upcoming vowel. If the formant values differ from one context to another, and if this difference occurs in the direction of the upcoming vowel, we can say that there is an anticipation of V_2 in V_1 .

Fig. 3 shows a clear effect of V_2 on V_1 for adults, for both V_1 vowels and in both formant dimensions. However, Fig. 4 does not show any clear effect of V_2 on V_1 for children. In order to quantify these observations, we conducted statistical analyses based on a General Linear Mixed-Effects Model (cf. section 2.4.3) in which V_1 , V_2 and Age group were independent factors and the speaker was a random factor. For the dependent variables F1 and F2 of vowel V_1 , the main effects of V_2 and V_1 were found to be statistically significant. Pair-wise post-hoc tests reveal that when $V_2 = /i/$, the average value of the z-scored F1 values decreases by 0.196 as compared to when $V_2 = /a/$, and that the normalized mean value of the z-scored F2 values increases by 0.636.

More interesting, the $V_2 \times$ Age group interaction was found to be significant. This result indicates that the direction and/or the magnitude of the main effects of V_2 are different across age groups. Pair-wise post-hoc tests reveal that for children as

compared to adults, the decrease in F1 associated with the change from $V_2 = /a/$ to $V_2 = /i/$ is reduced by 0.168. Similarly, the increase of F2 induced by $V_2 = /i/$ as compared to $V_2 = /a/$ is reduced for children by 0.405. In sum, it seems that for children, the effect of V_2 on V_1 is negligible for F1 and significantly smaller than for adults for F2.



Figure 3: Distributions, for all adults, of vowels /a/ and /ε/ as V_1 in the V_1 -C- V_2 sequences when V_2 is either /a/ (red) or /i/ (blue).

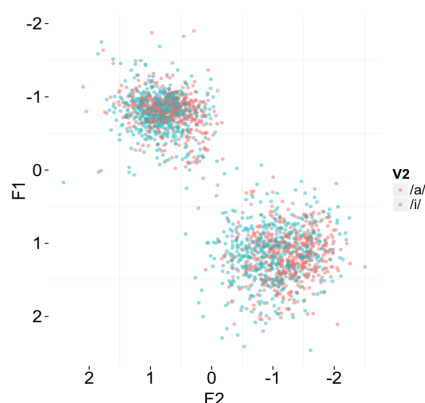


Figure 4: Distributions, for all children, of vowels /a/ and /ε/ as V_1 in the V_1 -C- V_2 sequences when V_2 is either /a/ (red) or /i/ (blue).

Figs. 5 to 7 display statistical representations of the sets of tongue contours measured at the midpoint of /ε/ as V_1 in /eda/ versus /edi/ sequences, respectively for an adult and for two children. These figures represent for the two different V_2 contexts, the average contours, together with the confidence intervals constructed using the SS ANOVA method (section 2.4.3). These examples illustrate well the fact that there is less token-to-token variability in adult's speech than in children's speech (see the smaller width of the confidence intervals for the adult speaker). The patterns observed in Fig. 6 correspond to those observed in most of our child speakers. It illustrates the fact that most of the children do not anticipate forthcoming gestures, since the confidence intervals calculated for both upcoming vowels V_2 largely overlap each other. However, for some of our child speakers, a small anticipation is observed in the antero-posterior dimension (see Fig. 7). For adults (Fig. 5), the effect of vowel V_2 is clear and significant along almost the entire tongue contour, with a clear anticipation in both dimensions and a strong consistency across trials.

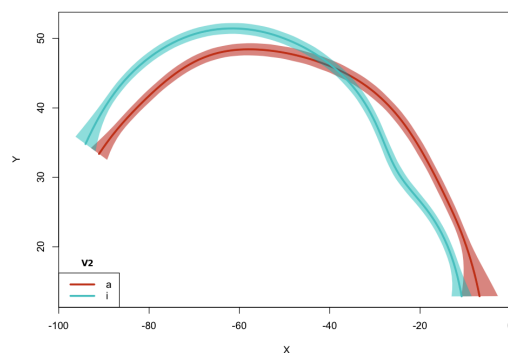


Figure 5: Statistical representations of the midsagittal tongue contours for /ε/ as V_1 in the contexts $V_2 = /a/$ (red) and $V_2 = /i/$ (blue) for an adult (Adult 6) speaker facing left.

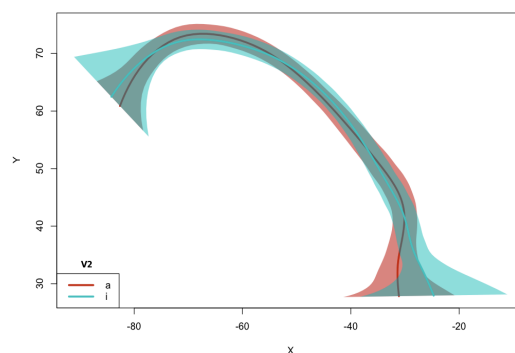


Figure 6: Same as figure 5 for child speaker 8.

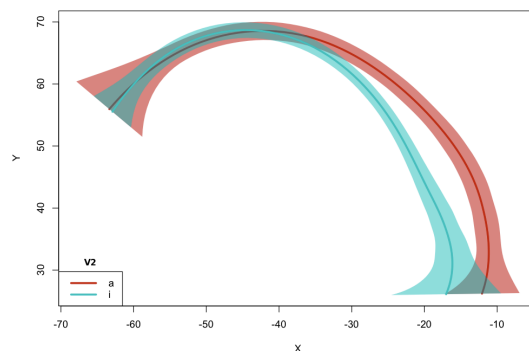


Figure 7: Same as figure 5 for child speaker 2.

4. Conclusion

Evidence has been found in the majority of the child participants for a greater token-to-token variability as compared to adults, and for an inability to anticipate V_2 in V_1 during the production of V_1 -C- V_2 sequences. In sum, these results indicate that 4-year-old children's speech motor control is immature from two perspectives: insufficiently stable motor control patterns for vowel production, and inability to plan speech sequencing by predicting the requirements of forthcoming gestures. Our interpretation supports the view that in 4-year-old children, the neural representations of the speech motor goals and of the speech motor system are immature.

5. Acknowledgements

This work is partly supported by FRQNT project N° 147877 (L. Ménard) and the project ANR-08-BLAN-0272 (P. Perrier).

6. References

- [1] Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M. K. and Perkell, J. S. (2013). Speech planning as an index of speech motor control maturity. In *Proceedings of Interspeech 2013*, Lyon, France, 1278-1282.
- [2] Lambert, J. and Bard, C. (2005). Acquisition of visuomanual skills and improvement of information processing capacities in 6- to 10-year-old children performing a 2D pointing task. *Neuroscience Letters*, 377, 1-6.
- [3] Jansen-Osmann, P., Richter, S., Konczak, J. and Kalveram, K.-T. (2002). Force adaptation transfers to untrained workspace regions in children: Evidence for developing inverse dynamic motor models. *Experimental Brain Research*, 143, 212-220.
- [4] Hourcade, J. P., Bederson, B. B., Druin, A. and Guimbretière, F. (2004). Differences in Pointing Task Performance Between Preschool Children and Adults Using Mice, *ACM Trans. Comput. Hum. Inter.* 11, 357-386.
- [5] Jansen-Osmann, P., Richter, S., Konczak, J. and Kalveram, K.-T. (2002). Force adaptation transfers to untrained workspace regions in children, *Experimental Brain Research*, 143, 212-220.
- [6] Walsh, B. and Smith, A. (2002). Articulatory movements in adolescents: evidence for protracted development of speech motor control processes. *Journal of Speech, Language and Hearing Research*, 45, 1119-1133.
- [7] Sadagopan, N. and Smith, A. (2008). Developmental Changes in the Effects of Utterance Length and Complexity on Speech Movement Variability. *Journal of Speech, Language and Hearing Research*, 51, 1138-1151.
- [8] Smith, A. and Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults, *Journal of Speech, Language and Hearing Research*, 41, 18-30.
- [9] Smith, A. (2010). Development of Neural Control of Orofacial Movements for Speech. In *Handbook of Phonetic Sciences*. W. Hardcastle and J. Laver (Eds). Oxford: Blackwell.
- [10] Zharkova, N., Hewlett, N. and Hardcastle W. J. (2011). Coarticulation as an Indicator of Speech Motor Control Development in Children: An Ultrasound Study. *Motor Control*, 15, 118-140.
- [11] Zharkova, N., Hewlett, N. and Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children, *Journal of the International Phonetic Association*. 42, 193-208.
- [12] Sereno, J. A. and Lieberman, P. (1987). Developmental aspects of lingual coarticulation. *Journal of Phonetics*, 15, 247-257.
- [13] Nittrouer, S., Studdert-Kennedy, M. and Neely, S.T. (1996). How children learn to organize their speech gestures: further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research*, 39, 379-389.
- [14] Siren, K.A. and Wilcox, K.A. (1995). Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech. *Journal of Speech and Hearing Research*, 38, 351-359.
- [15] Goodell, E. W. and Studdert-Kennedy, M. (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study. *Journal of Speech and Hearing Research*. 36, 707-727.
- [16] Noiray, A., Ménard, L. and Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America*, 133, 444-452.
- [17] MacNeilage, P. F. and Davis B. L. (2000). On the Origin of Internal Structure of Word Forms. *Science*, 288, 527-531.
- [18] Vihman, M. M. (1997). *Phonological development: the origins of language in the infant*. Blackwell, Oxford.
- [19] Whalen, D. H. (1990) Coarticulation is largely planned. *Journal of Phonetics* 18, 3-35.
- [20] Jordan, M.I. and Rumelhart, D.E. (1992). Forward models: supervised learning with a distal teacher, *Cognitive Science*, 16, 307-354.
- [21] Pollock, K. E. and Berni, M. C. (2003). Incidence of non-rhotic vowel errors in children: data from the Memphis Vowel Project. *Clinical Linguistics and Phonetics*, 17(4-5), 393-401.
- [22] Nazzi, T., Floccia, C., Moquet, B. and Butler, J. (2005). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, 102, 522-537.
- [23] Hallé, P. and Cristià, A. (2012). Global and detailed speech representations in early language acquisition. In S. Fuchs, M. Weirich, D. Pape and P. Perrier, (Eds.), *Speech production and perception: Planning and dynamics* (11-38). Frankfurt am Main: Peter Lang.
- [24] Bard, C., Hay, L. and Fleury, M. (1990). Timing and accuracy of visually directed movements in children : Control of direction and amplitude components. *Journal of Experimental Child Psychology*, 50, 102-118.
- [25] Forssberg, H., Eliasson, A. C., Kinoshita, H., Johansson, R.S. and Westling, G. (1991). Development of human precision grip I: Basic coordination of force. *Exp Brain Res*, 85, 451-457.
- [26] Forssberg, H., Kinoshita, H., Eliasson, A.C., Johansson, R.S., Westling, G., and Gordon, A. M. (1992). Development of human precision grip II: Anticipatory control of isometric forces targeted for object's weight. *Exp Brain Res*, 90, 393-398.
- [27] Vasudevan, E. V.L., Torres-Oviedo, G., Morton, S.M., Yang, J.F and Bastia, A.J. (2011). Younger Is Not Always Better: Development of Locomotor. *The Journal of Neuroscience*, 31(8), 3055-3065.
- [28] Perkell, J. S., Zandipour, M., Matthies, M. L. and Lane, H. (2002). Economy of effort in different speaking conditions I: a preliminary study of intersubject differences and modeling issues. *Journal of the Acoustical Society of America*, 112, 1627-1641.
- [29] Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E. and Hailey, D. S. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48, 543-553.
- [30] Boersma, P. and Weenink, D. (1996). Praat, a system for doing phonetics by computer, version 3.4, *Report No. 132, Institute of Phonetic Sciences of the University of Amsterdam*, 1-182.
- [31] Li, M., Kambhampettu, C. and Stone, M. (2005). Automatic contour tracking in ultrasound images, *Clinical Linguistics and Phonetics*, 6, 545-554.
- [32] Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York.
- [33] Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance, *Journal of the Acoustical Society of America*, 120, 407-415.
- [34] Bates, D., Maechler, M., Bolker, B. and Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen & S4*. R package version 1.0-5.
- [35] R Foundation for Statistical Computing. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- [36] Tremblay, A., University, D., & Ransijn, J. (2013). *LMERConvenienceFunctions: A suite of functions to back-fit fixed effects & forward-fit random effects, as well as other miscellaneous functions*, R package version 2.5.
- [37] Hothorn, T., Bretz, F. and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 3, 346-363.