



HAL
open science

Token-to-token variability and anticipatory coarticulation as indicators of maturity of speech motor control in 4-year-old children

Guillaume Barbier, Pascal Perrier, Lucie Ménard, Mark Tiede, Joseph Perkell

► To cite this version:

Guillaume Barbier, Pascal Perrier, Lucie Ménard, Mark Tiede, Joseph Perkell. Token-to-token variability and anticipatory coarticulation as indicators of maturity of speech motor control in 4-year-old children. ICA 2013 - Acoustics 2013 - 21st International Congress on Acoustics - 165th Meeting of the Acoustical Society of America, Jun 2013, Montréal, Canada. Paper 2aSC22. hal-00847607

HAL Id: hal-00847607

<https://hal.science/hal-00847607>

Submitted on 24 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013**

Speech Communication

Session 2aSC: Linking Perception and Production (Poster Session)

2aSC22. Token-to-token variability and anticipatory coarticulation as indicators of maturity of speech motor control in 4-year-old children

Guillaume Barbier*, Pascal Perrier, Lucie Menard, Mark Tiede and Joseph S. Perkell

*Corresponding author's address: Speech and Cognition, GIPSA-lab, Grenoble, 38000, Isère, France, guillaume.barbier@gipsa-lab.grenoble-inp.fr

Children's gestures do not appear to be executed with the same dexterity as adults'. Studies of arm movements have shown that young children's gestures are less accurate, more variable and slower than those of adults. This difference in behavior can be explained by a lack of experience with the sensory consequences of motor acts and still-developing forward models for the control of those acts. The hypothesis of immature and incomplete sensori-motor representations for speech in 4-year-old native speakers of Canadian French is addressed here through the analysis of ultrasound recordings of tongue contour kinematics and the speech signal from a corpus of isolated vowels and vowel-consonant-vowel sequences. Special attention is devoted to the analysis of vowel variability from two perspectives. Variability across repetitions in a single context provides information about the accuracy of the control. Contextual variability provides insights into the planning process as reflected in anticipatory coarticulation. Analysis of the observed lingual gestures will lead to improved understanding of the development of speech motor control and refinement of sensori-motor representations of speech. [Work supported by FQRNT project N 147877 and ANR project ANR-08-BLAN-0272]

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

The development of speech motor control has long been an important component of research on speech production and its underlying mechanisms. Numerous studies have investigated the abilities of children to produce speech, and have proposed and employed a variety of measures derived from the acoustic signal to quantify speech production skills. However, the development of speech production above the age of 2 (acquisition of the first word combinations), and more specifically the development of the spatio-temporal organization of speech gestures, remains poorly understood because of the difficulty of making direct measurements of children's articulatory movements during speech. In addition, contradictory experimental findings have been found, leading to various interpretations within different theoretical frameworks. Developmental studies of other aspects of human motor control, focused mainly on arm movements, have offered more consistent experimental findings that can be interpreted in a relatively unified theoretical framework. For speech, recent technological advances now support the measurement of children's tongue movements using ultrasound imaging. We believe that experimental studies using articulatory measures of children's speech, with findings interpreted in the more global framework of motor control, can lead to improved understanding of the development of speech motor control.

Numerous experiments on arm movements have investigated the ability of children to perform reaching or pointing tasks under various conditions. These studies (e.g. Jansen-Osmann et al., 2002 ; Hourcade et al., 2004; Lambert and Bard, 2005) have shown that children (aged 4-10 years) produce slower, less accurate and more variable movements than adults. This difference in behavior can be explained by a lack of experience with the sensory consequences of motor acts and still-developing forward models for the control of those acts. In other words, children do not have the same dexterity as adults because they do not have robust enough neural representations of their speech motor systems (internal models), particularly in terms of the amount of produced variability that is compatible with correct perception of the sound by listeners.

The aim of this study is to test this hypothesis in speech, by measuring two indices of the maturation of speech motor control in 4-year-old children and in adults: token-to-token variability and cross-syllabic anticipatory coarticulation. Reductions of token-to-token variability with age in children's speech production have been found in a number of studies (e.g. Nittrouer 1993, Smith & Goffman, 1998, Lee et al., 1999, Nijland et al., 2002, Walsh & Smith 2002, Nittrouer et al., 2005, Walsh et al., 2006). This indicates that the stability of the control increases with age. Another interesting index is in measures of anticipatory coarticulation, which is assumed to involve gesture planning. As such it seems to be an interesting indicator of the maturity of speech motor control. More specifically, cross-syllabic anticipatory coarticulation, influenced mainly by motor planning and less by the physical constraints of the speech articulators, can provide helpful information for understanding the development of speech motor control. The speech production system has an excess of degrees of freedom, i.e. many articulatory positions can be associated with perceptually similar sounds. Planning and producing a sequence of speech gestures consists of selecting gestures, for example according to an optimization criterion such as minimizing articulatory effort (a strategy inferred from measures in adults). This process requires implicitly learning about these degrees of freedom, which involves experience with the articulatory variability allowed for each phoneme in a speech sequence. In this sense, a measure of cross-syllabic anticipatory coarticulation in young children will provide information about the formation of their speech representations. To our knowledge, there is no developmental study in motor control focused on sequential tasks. Thus, it would be helpful to know how children plan and execute a sequence of gestures and to compare the control of arm and speech movements in the execution of sequential tasks.

Findings from the literature on anticipatory coarticulation in the speech of children are unfortunately quite contradictory: some authors have found that children exhibit less coarticulation than adults (e.g. Hodge, 1989); others found that children have the same patterns of coarticulation as adults (e.g. Sereno et al., 1987); and yet others have found that children exhibit a greater amount than adults (e.g. Nittrouer et al., 1996). The discrepancies in the results may be due to important differences in the nature of the objectives and tasks (intra-syllabic versus cross-syllabic coarticulation, nature of the phonemes used in the corpora, nature of the articulators involved in the anticipation, e.g., lips versus tongue), to the measures made (usually, only acoustic measurements, with

measurement time-points varying among studies) and to the complexity of the phenomenon, which is influenced both by gesture planning and biomechanical constraints of the articulators. The development of the spatio-temporal organization of speech gestures is poorly understood partly because almost all existing studies are limited to acoustic measurements. Because children's formants are difficult to track, and because acoustic measurements only provide indirect evidence of the underlying articulations, the understanding of speech motor control development could benefit from more direct articulatory evidence of children's behavior. The development of ultrasound tongue imaging and its use in developmental studies (e.g. Zharkova et al., 2011, 2012 ; Noiray et al., 2013) offers a promising method for the observation and measurement of children's tongue movements during speech.

To address the question whether 4-year-old typically developing children have immature and incomplete sensory-motor representations for speech, we have designed an experiment in which measurements of trans-syllabic coarticulation (vowel-to-vowel coarticulation in a VCV sequence) and token-to-token variability are obtained using simultaneous ultrasound tongue imaging and acoustic measurements of the speech signal. Our results will be interpreted in a theoretical framework coming from general motor control hypotheses and adapted to speech (Perrier, Loevenbruck, Payan, 1996 ; Perrier & Ma, 2008 ; Perrier, 2012). In this framework, anticipatory coarticulation is seen as a process in which optimal planning is performed. The experimental setup is primarily designed to reveal the extent to which 4-year-old children show immature patterns of planning of speech gestures.

METHODS

Participants

Twenty young Canadian French children aged 4 years old (4 years 0 months to 4 years 11 months) and 10 Canadian French adults (18-28 years old) were recruited in Montréal for the experiment. Canadian French was the first language of all participants. All children lived among 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. Each participant and participant's parent in the case of children were informed about all the procedures before the experiment and gave their consent. This study was approved by the ethical committee of the Université du Québec à Montréal (UQAM), as a part of a larger project investigating speech development (PI: Lucie Ménard).

Device

Among all devices used currently to record and measure tongue movements, few are appropriate for very young children. In contrast, ultrasound tongue imaging is a noninvasive exception, well-adapted for very young children, that allows a real-time view of most of the tongue, with good temporal (15Hz-90Hz) and a fair spatial (<1mm) resolution. To allow reliable measurements of tongue movements, a number of schemes have been proposed to solve the problem of head movement during ultrasound recordings. In this study, we used the HOCUS system (Haskins Optically Corrected Ultrasound System, Whalen et al., 2005), which uses the Optotrak (NDI Certus) tracking of infrared emitting diodes (IREDs) positioned both on the ultrasound probe and on the head of the participant to support representation of the data in a movement-corrected head-centric frame of reference. This approach is appropriate for developmental studies, in that it leaves freedom of movement to the participants.. A more detailed description of this system is given in Whalen et al., 2005.

In this study, we tracked the tongue in the midsagittal plane. Four IREDs are placed on the ultrasound probe; three are placed on the head of the participants (for head movement correction) and one is placed on the chin, on the midsagittal plane. This allows to tongue movements to be dissociated from jaw movements and provides an index of jaw motion.

Synchronous recordings of tongue movement in the midsagittal plane (at NTSC 29.97 Hz) and of the speech signal (at 44.1kHz) were made by the ultrasound device (Sonosite 180 Plus) and to a microphone both connected to a computer. The Optotrak system used in this study also recorded synchronously the audio and the positions of the

IREDs. A post-synchronization of these data is obtained by a cross-correlation between the two audio signals. After re-alignment, the data are presented in a 3D view in which the position of the IREDs and the plane in which the tongue image was taken are visible. Images taken too far from the midsagittal plane or with a significant angle to the midsagittal plane are removed and not used in the study.

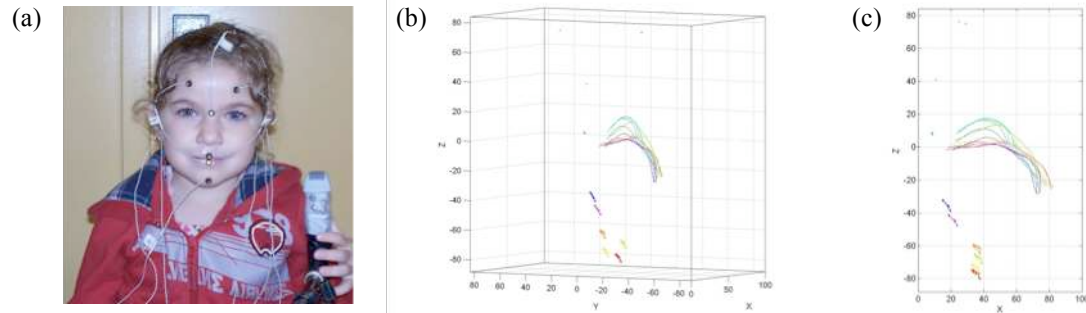


FIGURE 1. (a) A participant with the device installed (the ultrasound probe under the chin and IREDs, both on participant's head and on the ultrasound probe). (b) and (c) Re-aligned data corresponding to an /aga/ sequence, with tongue contours extracted for each frame, (b) in a 3D view and (c) in the midsagittal view, showing the position of the IREDs and the tongue contours during the sequence.

Task

The experiment was held at infant schools of Montréal and at the Laboratoire de Phonétique, UQAM. Participants were seated with the ultrasound probe under their chin, held by a microphone stand. They were seated in front of the Optotrak, which was also used in this experiment as a puppet theater. One experimenter checked that the head of the subject was not moving too much in reference to the ultrasound probe, and that most of the tongue was visible on the screen; another experimenter controlled the recording (Optotrak and ultrasound) via a computer and checked that all the IREDs were visible during the trials, and a third experimenter was the puppet master.

The task was presented as a puppet game. Puppets were presented in different pairs. The order of appearance of the pairs was randomized. The game took place as follows: the first puppet presents alone, and asked the participant to pronounce its name. The second puppet does the same. When the participant knows the names of the two puppets, the game starts. The participant must pronounce the name of the puppet when it appears. In this way, participants had to recall from memory, plan and execute a speech gesture or a sequence of speech gestures.

The corpus was collected as following, with 8 to 10 repetitions for each isolated vowel or VCV sequence:

- Isolated vowels /i e ε a u/
- Vowel-Consonant-Vowel /V₁-C-V₂/ sequences with C = /b d g/, V₁ = /ε a/ and V₂ = /i ε a/

The isolated vowels are used to measure the dispersion in F1-F2 space of vowel categories and to measure the token-to-token variability of these vowels. They are also used to ensure that 4-year-old children produced differentiated /e/ and /ε/. The /V₁-C-V₂/ sequences are composed of non-constrained vowels as V₁ (a certain variability in the production of the vowels /a/ and /ε/ is permitted), and with a high vowel /i/ and a low vowel /a/ as V₂, as well as V₁=V₂ as the reference sequence. These sequences were designed to measure the effect of anticipating V₂ in measuring V₁ realizations.

Data Analysis

The acoustic data were first labeled with *Praat* (Boersma & Weenink, 1996). For vowels, the beginning of the vowel is defined as the first descending zero-crossing, and the end of the vowel is defined as the time when F2 disappears. For the stops, the beginning of the consonant is defined as the time where F2 of the preceding vowel disappears and the end corresponds to the beginning of the release burst. The beginning and the end of the burst noise are also labeled. The transition from the stop to the subsequent vowel starts at the noise onset and, ends at the moment when F2 appears.

The acoustic signal was down sampled to 22050 Hz in order to achieve more accurate formant detection. Automatic acoustic measurements of the formants in the midpoint of the vowels are made with a Linear Predictive Coding (LPC) algorithm. Because formant tracking is complex in child speech, and can lead to detection errors, we compared maxima detected in the LPC spectra with the frequencies of the poles in the LPC filter. A range of possible formants values for each vowel was specified beforehand, to remove outliers.

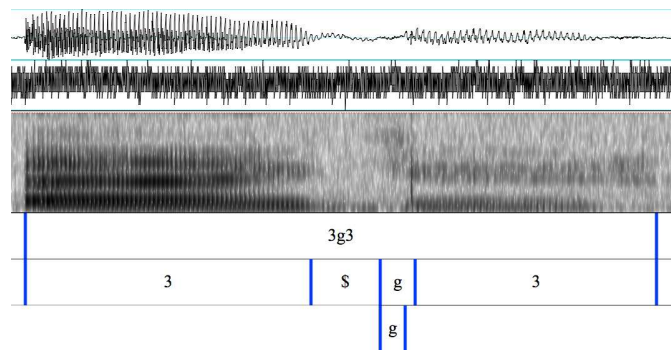


FIGURE 2. Example of tiers and boundaries used in *Praat* to label the acoustic data. The first tier corresponds to the whole VCV sequence. The second corresponds to the duration of each segment, and is used to measure segment durations and vowel formants. The last one is only devoted to measure the stop noise.

Concerning the articulatory data, the US images closest to the midpoint of the vowels were extracted. The upper tongue contour was extracted using a semi-automatic procedure developed for the purpose, *GetContour*. It is similar to other semi-automatic tongue contour extraction software such as *EdgeTrak* (Li et al., 2005) in many ways, but it also allows flexible tracking of the contour of one image alone, which is useful for the present study.

Contours were converted to 3D head-centric coordinates using the HOCUS procedures described above.

RESULTS

As data analysis is currently in progress, results will be presented at the conference. Below are the expected results.

First, we expect to find more token-to-token variability in children than in adults, and to find more within-subject variability and between-subjects variability in children, reflecting the lack of accuracy of the motor control system for speech production in children and the differences of maturity between them and adults. Token-to-token variability, measured as the F1-F2 dispersion of an isolated vowel or a vowel in context and in terms of distance (mm) between the tongue contours, among 8-10 repetitions, is expected to be significantly greater in children than in adults. This prediction is made according to the hypothesis that 4-year-old children have a lack of experience with the sensory consequences of motor acts and still-developing forward models for the control of those acts. This will lead to reduced accuracy of the control for the production of a single phoneme.

Second, we expect to find less vowel-to-vowel anticipatory coarticulation in children than in adults. Immature and incomplete sensory-motor representations for speech will be associated to a lesser influence of the context in the

production of a sequence of speech gestures. In our view, adults will exhibit very stable and optimal patterns of coarticulation in the sense that they will always use the pattern that allows the minimum of motor command changes (the minimum of effort) during the production of the speech sequence. This behavior requires good enough representations of the speech motor system in terms of allowed variability. Children will exhibit more variable patterns of coarticulation from one trial to another and will exhibit less optimal patterns of coarticulation. Adults will show greater anticipation of V_2 in V_1 . We expect to measure significantly higher V_1 in the / V_1 -C-i/ context and significantly lower V_1 in the / V_1 -C-a/ context compared to the reference context of $V_1=V_2$. We expect that this behavior will be not observed in children, for whom the differences between these contexts will be not statistically significant.

In addition to these two major expected results, we also expect to measure a greater amount of coarticulation with C as a bilabial than in all other contexts, less with C as a velar, and less with C as an alveolar. These patterns are also expected to be found in children, with less global effect of the consonantal context on vowel realizations, according to the hypothesis of immature and incomplete sensory-motor representations for speech. According to the hypothesis that children have still-developing forward models for the control of speech gestures, we also expect to find that children produce slower movements than adults. Indeed in the absence of predicting forward models, the children have to rely on auditory or oro-sensory feedback to monitor their speech. These feedback loops last for more than 100 ms.

CONCLUSION

This study is aimed at understanding the development of speech motor control and characterizing the maturity of sensory-motor representations for speech in 4-year-old children, through the analysis of token-to-token variability and vowel-to-vowel anticipatory coarticulation. If the expected results are confirmed, they will shed new light on the understanding of the refinement of sensory-motor representations of speech. Our results can be interpreted in a framework coming from a literature review of motor control and motor control development studies, adapted to speech. In this framework, the sensory consequences of motor acts are learned and stored in forward models, and further used to predict the sensory consequences of those acts. This framework is tested here through the development of speech production.

ACKNOWLEDGMENTS

This work is partly supported by FQRNT project N° 147877 (Lucie Ménard) and ANR project ANR-08-BLAN-0272 (Pascal Perrier). We want to thank Amélie Prémont, Christine Turgeon, Paméla Trudeau-Fisette and Dominique Côté for their crucial help concerning the experiment. We are thankful to the infant school Le Sablier and particularly its co-director Carole Langlois for her involvement in the project. We also thank all our participants.

REFERENCES

- 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.
- Hodge, M. (1989). "A comparison of spectral-temporal measures across speaker age: Implications for an acoustic characterization of speech maturation," Unpublished doctoral dissertation, University of Wisconsin-Madison.
- 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**. 4. 357-386.
- Jansen-Osmann, P., Richter, S., Konczak, J. and Kalveram, K.-T. (2002). "Force adaptation transfers to untrained workspace regions in children," *Exp. Brain Res.* **143**, 212–220.
- 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.

- Lee, S., Potamianos, A., and Narayanan, S. (1999). "Acoustics of children's speech: developmental changes of temporal and spectral parameters," *J. Acoust. Soc. Am.* , **105**, 1455–1468.
- Li, M., Kambhampettu, C., and Stone, M. (2005). "Automatic contour tracking in ultrasound images," *Clin. Ling. Phonetics* **6** (19), 545–554.
- Nijland, L., Maassen, B., Van der Meulen, S., Gabrieëls, F., Kraaimaat, F.W., and Schreuder, R. (2002). "Coarticulation patterns in children with developmental apraxia of speech," *Clinical Linguistics & Phonetics*, **16**, 461–483.
- Nittrouer, S. (1993). "The emergence of mature gestural patterns is not uniform: evidence from an acoustic study," *Journal of Speech and Hearing Research*, **36**, 959–972.
- 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.
- Nittrouer, S., Estee, S., Lowenstein, J.H., and Smith, J. (2005). "The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops," *J. Acoust. Soc. Am.* , **117**, 351–364.
- Noiray, A., Ménard, L. and Iskarous, K. (2013) "The development of motor synergies in children: Ultrasound and acoustic measurements," *J. Acoust. Soc. Am.* **133** (1), 444-452.
- Perrier, P., Lœvenbruck, H. and Payan, Y. (1996) "Control of tongue movements in speech: The Equilibrium Point hypothesis perspective," *Journal of Phonetics*, **24**, 53-75.
- Perrier, P. and Ma, L. (2008). "Speech Planning for VCV Sequences: Influence of the Planned Sequence," in *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, France.
- Perrier, P. (2012) "Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control," in *Speech Planning and Dynamics*, S. Fuchs, D. Pape and P. Perrier (Eds.) 191-238.
- Sereno, J. A., Baum, S. R., Marean, G. C., and Lieberman, P. (1987). "Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children," *J. Acoust. Soc. Am.* **81**, 512–519.
- 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.
- 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.
- Walsh, B., Smith, A., and Weber-Fox, C. (2006). "Short-term plasticity in children's speech motor systems," *Developmental Psychobiology*, **48**, 660–674.
- Whalen, D. H., Iskarous, K., Tiede, M., Ostry, D. J., Lehnert-Lehouillier, H., Vatikiotis-Bateson, E. et Hailey, D. (2005) "The Haskins optically corrected ultrasound system (HOCUS)," *Journal of Speech, Language, and Hearing Research*, **48** (3), 543-553.
- Zharkova, N., Hewlett, N., and Hardcastle, W. J. (2011). "Coarticulation as an indicator of speech motor control development in children: An ultra-sound study," *Motor Control* **15**, 118–140.
- 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," *J. Int. Phonetic Assoc.* **42** (2), 193–208.