Phonetic Structure and Acquisition of Laryngeal and Pharyngeal Articulations: Text-Analysis Considerations

John H. Esling, Lisa D. Bettany, Allison Benner, and
Spencer Rose
University of Victoria
esling@uvic.ca; abenner@uvic.ca

Abstract

This article describes the development of an auditory coding system and an associated web-based XML database to document and analyze speech development in infants from Canada, Morocco, France, and China. The XML database will allow researchers on this international research project, based at the Department of Linguistics, University of Victoria, to collaborate in data collection and analysis, facilitating a better understanding of phonetic development in infants from diverse language backgrounds. We hypothesize that infants from all language backgrounds employ laryngeal constriction as the primary means of phonetic development in the first months of life, using this experience as a base to learn the sounds of their native language in the second half of the first year. Preliminary results from English-speaking infants support this hypothesis.

KEYWORDS: Speech acquisition, phonetics, laryngeal constriction, auditory coding, XML database.

Introduction

This article describes an international research project on infant speech acquisition based at the Department of Linguistics, University of Victoria, with particular emphasis on developing an auditory coding system and an associated web-based XML database to document and analyze the speech development process.

In recent years, considerable research has focused on infant vocalizations as precursors to communication and language development (Kent & Murray, 1982; Koopmans-van Beinum & van der Stelt, 1986; Oller, 1980, 2000; Stark, 1980; Stark et al., 1975). The primary goal of most previous research has been to demonstrate the progression of

vocalizations towards meaningful speech. Since the majority of researchers in this field have been speakers of Indo-European languages, most commonly English, they have tended to focus exclusively on "speech-like" sounds with "normal" or modal phonation from an English perspective, excluding laryngeally "strained" and "tense" vocalizations from their analyses (Koopmans-van Beinum & van der Stelt, 1986; Oller, 1980; 2000), on the grounds that these sounds are "vegetative" or "reflexive". However, given that these sounds comprise a large percentage of the vocalizations produced in early infancy (Esling, Benner & Bettany, 2004a; Bettany, 2004; Stark et al., 1975) and that they occur in a variety of languages spoken around the world, we believe this approach is unwarranted.

Furthermore, even when researchers do include laryngeal sounds in their studies (McCune et al., 1996; Stark et al., 1975), they usually describe them in colloquial terms such as "grunting", "coughing", groaning", "moaning", or "croaking". These impressionistic terms, all of which imply engagement of the laryngeal sphincter, do not provide an accurate description of the productive capability of infants or the integral role of laryngeal constriction in the vocalizations of infants. Furthermore, they limit opportunities to track infant speech development in a principled manner. While refined control of the structures required for many oral articulations (e.g., the muscles of the tongue) is underdeveloped in early infancy, efficient control of the primary laryngeal mechanism involved in voicing and breath (i.e. vocal fold vibration and opening) and protection of the airway (i.e. the laryngeal constrictor mechanism) is present at birth. We hypothesize that infants in all language environments use these efficiently coordinated mechanisms to explore the productive capacity of their voices, and ultimately, the phonetic features of their native languages.

We are now collaborating with researchers in France, China, and Morocco to document the speech development of infants from four different language backgrounds: English, French, Bai, and Moroccan Arabic. Two of these languages (Bai and Moroccan Arabic) employ laryngeal constriction within their linguistic systems, while the other two (English and French) do not. Nevertheless, we expect that infants from all four language backgrounds will make extensive use of laryngeal constriction in the first several months of life. We anticipate that this exploration will take a different form in the second half of the first year of life, depending on whether the infants' ambient language employs laryngeal constriction or not. To test this hypothesis, we have developed an auditory coding method that will allow us to document the speech development of infants

from these four language backgrounds in a phonetically principled manner. This coding method serves as the basis for a web-based XML database that will facilitate collaborative annotation and sharing of data and analysis over the course of the project.

1. Auditory Coding of Infant Articulations

The question of how to categorize prelinguistic vocalizations accurately has been a source of controversy among infant speech researchers (Nathani & Oller, 2001). Recently, Amano et al. (2003) and Buder & Oller (2003) have suggested that auditory-based coding utilizing a standardized database model of infant vocalizations is more accurate than previous strategies, including acoustic analysis (Lieberman, 1985), and articulatory or phonatory description (Koopmans-van Beinum & Van der Stelt, 1986; Stark, 1980). Training researchers with access to specific examples located in user-friendly databases may provide more accurate and synchronous descriptions of infant vocalizations.

As the basis for our web-based auditory coding method, we have developed an auditory-phonetic model of the larynx to account for the infants' exploration of constricted laryngeal phonation in the first year of life. Based on laryngoscopic observations from a large sampling of languages<http://web.uvic.ca/ling/research/phonetics/index.htm>, Esling and colleagues (Edmondson et al., 2001; Esling, 1996, 1999a, 1999b, 2002, 2003; Esling & Edmondson, 2002; Esling & Harris, 2003) have developed canonical profiles of adult laryngeal productions. In this study, this model of the larynx was applied to infant vocalizations to describe phonetically the productive capability of the laryngeal mechanism in the first year of life.

Our method condenses the complexity of existing colloquial terms into two main auditory categories: constricted (i.e. involving the engagement of the vertical laryngeal constrictor) and unconstricted (i.e. involving the horizontal level of the glottis). These two primary categories can be broken down into more precise and descriptive auditory categories. For example, constricted settings include whisper, harsh voice, and creaky voice; and unconstricted settings include modal voice and falsetto. These categories can be further subdivided to capture a range of auditory phonetic parameters, including degree of constriction, presence or absence of voicing, and pitch level (high, mid, or low, according to predefined criteria regarding the source and frequency of vibrations).

To classify short-term utterances that are roughly equivalent to adult segmental articulations, we are using the categories for place and manner of articulation provided by the International Phonetic Alphabet. The physical maneuvers involved in "grunts", for example, are more accurately described in phonetic terms as voiced or voiceless pharyngeal fricatives or trills, while "coughs" usually correspond to voiced or voiceless pharyngeal stops.

This auditory coding method allows project researchers to classify the full range of sounds that infants make, providing a different interpretation of the origins of infant phonetic development than those found in previous studies. The auditory categories also provide the basis for a metadata editor and a searchable, web-based XML database.

2. XML Database

The Internet has created opportunities for researchers to share data worldwide. In making our infant speech data available on the Web, we join a number of other prominent researchers in this field who are developing comprehensive databases of infant vocalizations, including Lorraine McCune (Rutgers University), Kim Oller (University of Maine), Marilyn Vihman (University of Wales), Eugene Buder (University of Memphis), and Shigeaki Amano (NTT Communication Science Laboratories). These databases expand possibilities for researchers to develop universal theories to account for infants' phonetic development. Among these databases, ours is unique in including both audio and video files illustrating the full range of infant vocalizations, and providing a sample of infant vocalizations produced in four very different language environments.

With the assistance of the Humanities Computing and Media Centre at the University of Victoria, we have developed a metadata editor which we use to record information on the infants in our study and the vocalizations they produce. Once entered, the metadata forms part of an XML file that is searchable via a Web-based interface,<http:lingo.dcf. uvic.ca.8080/exist/pisa> (see Figure 1 below for a sample view of the Web interface).
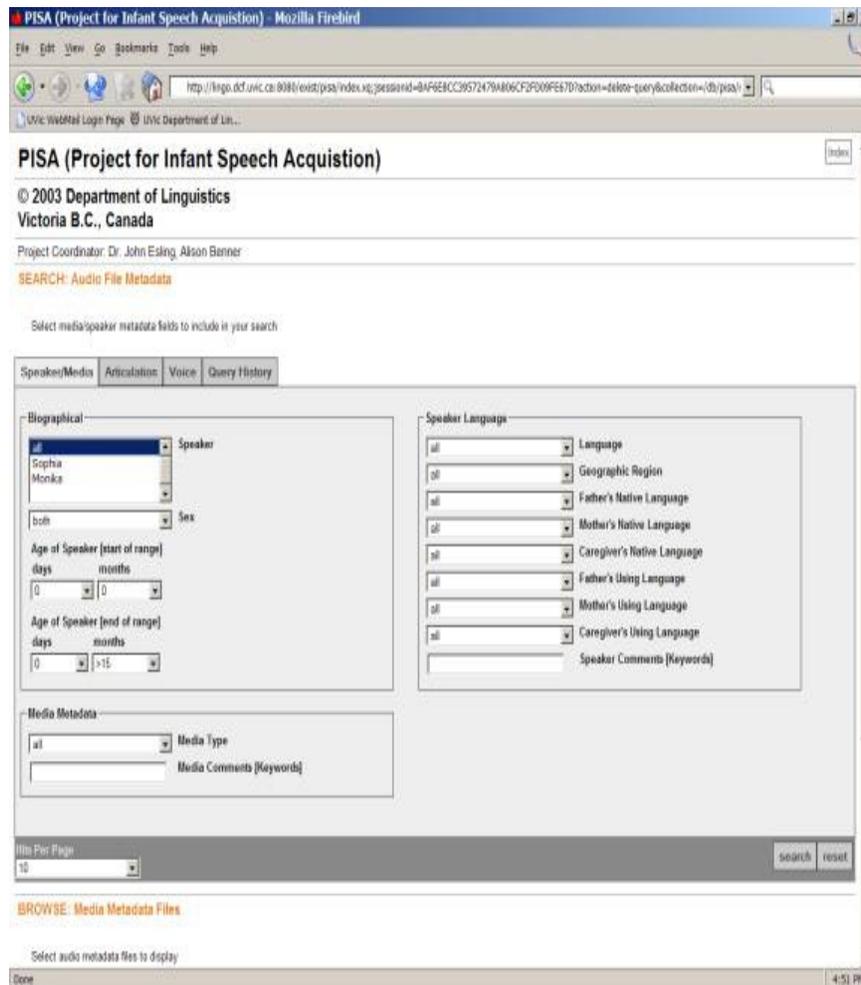
Figure 1: Web Interface of the Project for Infant Speech Acquisition

This application uses an implementation of the Java-based XML database system eXist, which operates both as a repository and search engine for XML documents. Unlike standard relational databases, an XML database preserves and utilizes the hierarchical nodal structure of XML to carry out sophisticated querying of the data, and return results as XML. As XML, the data can be saved and repurposed for future applications.

The eXist system includes an extensive library of extension modules that handle HTTP requests and database manipulation, and, importantly for this project, standard XML technologies of XQuery and XSLT.

Extensive support of XQuery and XSLT simplified the development of the metadata search interface: XQuery was used to create complex scripts for searching and filtering metadata; XSLT was used to deliver a useful interface. As a standard XML technology, XQuery is interoperable with other XML technologies, such as XPath, which use the same functions, operations, and data model. XPath, a language used to describe document nodal relations, is used to encode search criteria as a structured query. For most XPath queries, the eXist search engine refers to stored index files generated automatically when the file is added to the database collection. The use of these indexes significantly improves the speed and efficiency of XML database searches.

In the processing of queries for this application, search criteria submitted by users via standard HTTP requests are translated in XQuery into standard XPath expressions, which are then applied to the XML metadata. Given that, for this project, search criteria cover data fields from separate database collections (i.e. for speech metadata and speaker profile information), the XQuery search script is designed to cross-index these collections, and allow for a one-step searching and filtering of results. Also, given the extensive range of search criteria offered, a further feature of the interface is a "query history". This allows researchers to store and date search criteria selections that can be later re-applied to the database.

The XML database allows researchers to locate prototypical examples of various constricted and unconstricted infant articulations; and to perform searches by child, sex, age, language background, file type (audio or video), and auditory parameter. We currently have a growing database of articulations produced by infants in English-speaking environments in Canada, and will soon be adding articulations produced by infants from Morocco, China, and France. As our database grows, we will be in a position to produce fine-grained cross-linguistic comparisons of speech development of infants from these four language backgrounds. Researchers in all four locations can then begin to draw on the database to produce collaborative annotations of infant articulations. These annotations, along with spectograms and pitch contours, can then be made available for viewing via the search function, increasing opportunities for dialogue among researchers with shared interests around the world. Furthermore, as the project develops, the annotations themselves may provide a new source of searchable metadata, yielding a more nuanced understanding of infant speech development than was possible at the outset of the project.

3. Application of the Model

To date, we have used our auditory coding method and the associated database to analyze speech development in English-speaking infants. Overall, we have found that short-term segmental and longer-term phonatory articulations are primarily laryngeally constricted in the first few months of life, but become progressively less constricted towards the end of the first year of life (Benner et al., 2004; Esling et al. 2004a; 2004b). Given the characteristics of Canadian English, this is the pattern we would predict. For example, very young infants produce a high proportion of pharyngeal and glottal articulations, but these types of sounds do not occur in their ambient language. In the second half of the first year, the proportion of pharyngeal articulations declines, while alveolar, velar, and bilabial segments, which occur in English, steadily increase. Similarly, newborns of English-speaking parents produce a high incidence of highly constricted harsh voice−a voice quality that is not particularly common among speakers of Canadian English. By the end of the first year, these same infants produce a high proportion of unconstricted vocalizations, including modal voice, falsetto, and breathy voice−qualities that predominate in their speech community. These findings are consistent with our hypothesis that infants use a constricted setting, to which they are biologically predisposed, to explore their voices and, eventually, the phonetic features of their native languages.

Drawing on the methods outlined above, Bettany (2004), a member of our phonetics research team at the University of Victoria, has further refined our understanding of infant speech development in her study of speech development in the first six months of life. Her analysis of 824 vocalizations produced by a Canadian English infant provides a sample of the type of analysis we will be able to extend to our study of other Canadian infants, as well as infants from France, Morocco, and China. In line with our previous findings, Bettany found that the infant's vocalizations were predominantly constricted in the first months of life, but grew increasingly unconstricted towards the end of the first six months, with the fifth month marking an important transition for this particular infant, as shown in Figure 2 below.
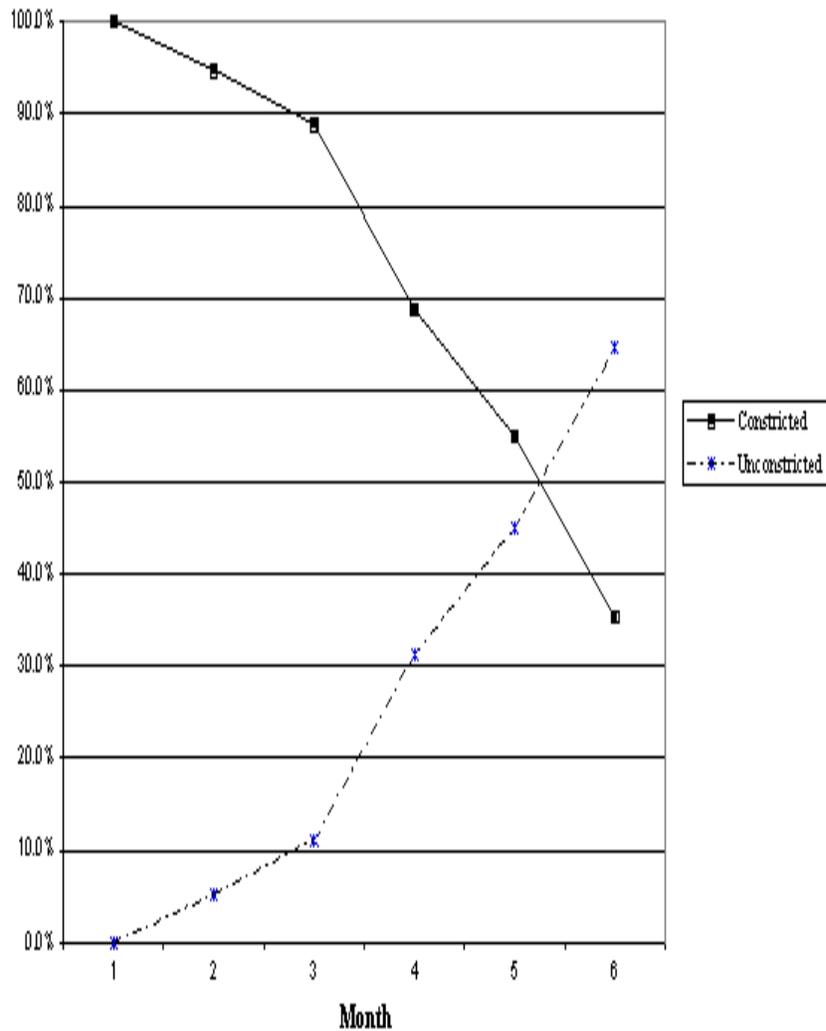
Figure 2. Percentage of constricted and unconstricted vocalizations in the
first six months of life. (Bettany 2004: 60)

Bettany added to our team's findings by conducting a detailed analysis
of the interaction of pitch and constriction and, in turn, of alternations
between different types of vocalizations−a type of analysis that is very
fruitful in the understanding of speech development and that will be
extended to a larger and more diverse group of infants as our infant speech

database grows.   In the speech development of this particular infant, the fifth month of life was a critical period not only in the transition from predominantly contricted to unconstricted settings, but in the exploration of pitch and an increase in the incidence of alternations between constricted and unconstricted settings at various pitch levels. In Bettany's analysis, the expanding repertoire of the infant seems to be occurring primarily as a function of increased laryngeal control, highlighting the importance of the laryngeal sphincter in speech development. As we extend our analysis to French, Moroccan Arabic, and Bai-speaking Chinese infants, we will be interested to see how infants from a variety of backgrounds exploit the same biological tendencies and associated phonetic parameters to reach different linguistic outcomes.

Conclusion

This paper has provided a brief overview of our collaboration with other researchers around the world to document infant speech development. We believe that our web-based XML database will allow researchers to share, exchange, and compare a large amount infant vocalization data in ways that were not possible even in the recent past. This collaboration, still in its first stages, will enhance our understanding of the role of laryngeal mechanisms in the phonetic development of infants in the first year of life.

Works Consulted

Amano, S., T. Nakatani and T. Kondo. (2003). "Fundamental Frequency Analysis of Longitudinal Recording in a Japanese Infant Speech Database." *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* 2: 12-518.

Benner, A., L. Bettany and Q. Wang. (2004). "Natural Language Processing in the First Year of Life." Poster presentation, Advanced Systems Institute Exchange, Vancouver, March 8, 2004.

Bettany, L. (2004). "Range Exploration of Phonation and Pitch in the First Six Months of Life." M.A. thesis, Department of Linguistics, University of Victoria.

Buder, E. H., K.D. Oller and J.C. Magoon. (2003). "Vocal Intensity and Phonatory Regimes in the Development of Infant Protophones." *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* 3: 2014-2018.

Edmondson, J.A., Lama Ziwo, J.H. Esling, J.G. Harris and Li Shaoni. (2001). "The Aryepiglottic Folds and Voice Quality in the Yi and Bai Languages: Laryngoscopic Case Studies." *Mon-Khmer Studies* 31: 83-100.

Esling, J.H. (1996). "Pharyngeal Consonants and the Aryepiglottic Sphincter." *Journal of the International Phonetic Association* 26: 65-88.

---. (1999a). "The IPA Categories 'Pharyngeal' and 'Epiglottal': Laryngoscopic Observations of Pharyngeal Articulations and Larynx Height." *Language and Speech* 42: 349-372.

---. (1999b). "Voice Quality Settings of the Pharynx and Larynx." *Proceedings of the 14th International Congress of Phonetic Sciences*, Vol. 3. 2449-2452.

---. (2002). "Laryngoscopic Analysis of Tibetan Chanting Modes and their Relationship to Register in Sino-Tibetan." Proceedings of the 7th International Conference on Spoken Language Processing, Vol. 2. 1081-1084.

---. (2003). "Glottal and Epiglottal Stop in Wakashan, Salish and Semitic." *Proceedings of the 15th International Congress of Phonetic Sciences*, Vol. 1. 1707-1710.

---. A. Benner, and L. Bettany. (2004a). "Phonetic Articulatory Control in the First Year of Life." Paper presented at the 2004 Conference of the British Association of Academic Phoneticians, Cambridge University, March 24-26, 2004.

---. A. Benner, L. Bettany and C. Zeroual. (2004b). "Le contrôle articulatoire phonétique dans le prébabillage." *Actes des XXVes Journées d'Etude sur la Parole, 19-22 avril 2004, Fès, Maroc.* Aix-en-Provence: Association Francophone de la Communication Parlée.

---. J.A. Edmondson. (2002). "The Laryngeal Sphincter as an Articulator: Tenseness, Tongue Root and Phonation in Yi and Bai." *Phonetics and*

its Applications: Festschrift for Jens-Peter Kőster on the Occasion of his 60th Birthday. Eds. A. Braun and H.R. Masthoff. Stuttgart: Franz Steiner. 38-51.

---. J.G. Harris. (2003). "An Expanded Taxonomy of States of the Glottis." *Proceedings of the 15th International Congress of Phonetic Sciences,* Vol. 1. 1049-1052.

Kent, R.D. and A.D. Murray. (1982). "Acoustic Features of Infant Vocalic Utterances at Three, Six, and Nine Months." *Journal of the Acoustical Society of America* 72: 353-365.

Koopmans-van Beinum, F.J. and J.M. van der Stelt. (1986). "Early Stages in the Development of Speech Movements." *Precursors of Early Speech.* Eds. B. Lindblom and R. Zetterstrőm. New York: Stockton. 37-50.

Lieberman, P. (1985). "The Physiology of Cry and Speech in Relation to Linguistic Behavior." *Infant Crying: Theoretical and Research Perspectives*. Eds. B. Lester and C. F. Z. Boukydis. New York: Pleum Press. 37-52.

McCune, L., M.M. Vihman, L. Roug-Hellichius, D. Bordeneave, and L. Gogate. (1996). "Grunt Communication in Human Infants (homo sapiens)." *Journal of Comparative Psychology* 110: 27-37.

Nathani, S. and D.K. Oller. (2001). "Beyond ba-ba and gu-gu: Challenges and Strategies in Coding Infant Vocalizations." *Behavior Research Methods, Instruments, & Computers* 33: 321-330.

Oller, D. K. (1980). "The Emergence of the Sounds of Speech in Infancy." *Child Phonology*. Eds. G. Yeni-Komshian, J., Kavanaugh, & C. Ferguson. New York: Academic Press. 93-112.

Oller, K. D. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Association.

Stark, R.E. (1980). "Prespeech Segmental Feature Development." *Child Phonology*. Eds. G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson. New York: Academic Press. 73-92.

---. S.N. Rose and M. McLagen. (1975). "Features of Infant Sounds: The First Eight Weeks of Life." *Journal of Child Language* 2: 205-221.