Acoust. Sci. & Tech. **33**, 4 (2012)

INVITED REVIEW

Development of vocal tract and acoustic features in children

Ryoko Mugitani* and Sadao Hiroya

NTT Communication Science Laboratories, NTT Corporation,
3–1, Morinosato Wakamiya Atsugi, 243–0198 Japan

Keywords: Child development, Vocal tract, Acoustic change, Formant

PACS number: 43.70.Ep [doi:10.1250/ast.33.215]

1. INTRODUCTION

When we vocalize, the vibration of the vocal folds becomes a primal sound source and the shape of the vocal tract from the glottis to the lips determines the resonance of speech sounds, which specifies vowels and consonants. A considerable number of studies have already explored the anatomical and acoustic characteristics of adult vocal tracts, but only a few studies have attempted this in relation to the developmental changes from infants to children to adults. It is well known that an infant's vocal tract is not simply a miniature version of an adult's vocal tract. Therefore, the developmental trajectory will not be uniform and each segment of the vocal tract is expected to grow in a different manner and with different timing.

This article reviews recent research on vocal tract development. In the first half of this paper, we illustrate the anatomical development of the vocal tract that has been revealed by using computed tomography (CT) and magnetic resonance imaging (MRI). The anatomical development of the vocal tract, which is a resonant tube, is accompanied by a change in acoustic features. Therefore, in the second half of the article we introduce studies that focus on the acoustic features of a child's utterance and its developmental change.

2. ANATOMICAL DEVELOPMENT OF VOCAL TRACT

It is well known that the vocal tract of a newborn infant is rather closer to that of a primate than that of an adult human being [1] (Fig. 1). The simplest difference between a child's and an adult's vocal tract is the length. The vocal tract lengthens from about 8 cm in newborns to about 17 cm in adults. In addition, the bend in the oropharyngeal channel is gradual at the beginning of life. The soft palate and the epiglottis, which prevents food and drink from entering the airway during swallowing, are very close to each other. During postnatal development, the hyoid bone and larynx descend gradually and the bend in the oropharyngeal channel almost forms a right angle. As a result of the overall postnatal development, the ratio of the vertical pharynx length versus the horizontal oral cavity length changes from 1:2 at the beginning to approximately 1:1 in adulthood [2].

The anatomical structure of the vocal tract had traditionally been studied by radiographic techniques. Therefore, the risk of radiation exposure had limited the subjects exclusively to adults (but see as exception [3]). However, the recent appearance of several imaging techniques applicable for children, e.g. the widely used MRI, has led to studies of anatomical morphology and of vocal tract development. Another advantage of MRI is that it provides clear images of soft tissue such as the velum and lips. (See [4] for more general information about anatomical and acoustic studies using MRI.)

Since the year 2000, Vorperian, Kent and their colleagues at the University of Wisconsin-Madison have conducted a series of studies using a large-scale dataset, i.e. 605 MRI and CT images of individuals between birth and 19 years, to intensively explore vocal tract development [5–7]. In their studies, nine variables that represent the lengths of nine segments of a vocal tract (Fig. 2) were measured to determine the growth pattern and speed of the vocal tract. For example, they performed various polynomial model fits to obtain the growth curves of the vocal tract, and found that the fourth degree model was the best fit to the data [6]. This fact indicates the non-uniform and non-linear growth pattern of the vocal tract and that the growth speed varies depending on the phase of development. In particular, the growth of all nine segments is remarkably rapid in the first few years of life so that the vocal tract lengthens about 2 cm during the first 2 years. The rapid descent of the larynx and the hyoid bone has also been reported for the same time period [3,5].

The recent studies also revealed that the growth speed differs depending on the segments of the vocal tract. For

*e-mail: mugitani.ryoko@lab.ntt.co.jp

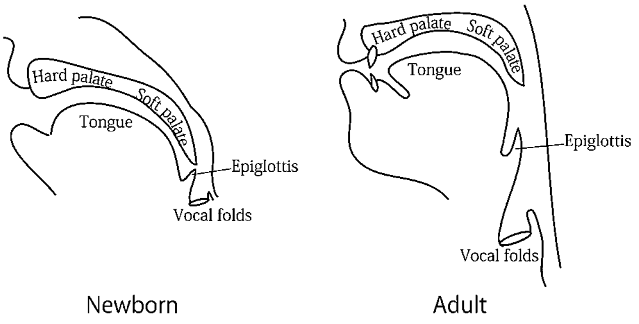

Fig. 1 Vocal tracts of newborn infant and adult.

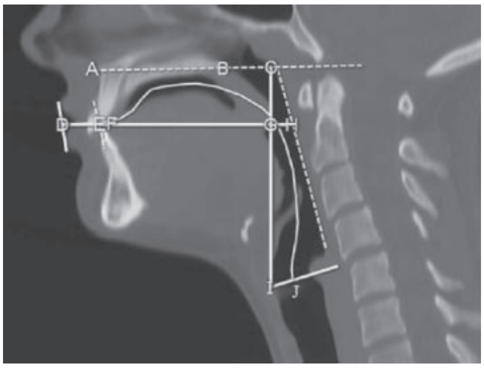

Fig. 2 Midsagittal vocal tract image displaying the anatomic landmarks used in [6,7]. The nine variables studied are as follows. Variable (1) is vocal tract length (VTL), which is the curvilinear line extending from points D to J. Variable (2) is vocal tract-vertical (VT-V), which is the vertical distance from points I to C. VT-V consists of two segments (which are Variables 3 and 4, respectively): (3) posterior cavity length (PCL; points I to G) and (4) nasopharyngeal length (NPhL; points G to C). Variable (5) is vocal tract-horizontal (VT-H), which is the horizontal distance from points D to H. VT-H consists of three line segments, which are Variables 6, 7, and 8, respectively: (6) lip thickness (LTh; points D to E), (7) anterior cavity length (ACL; points F to G), and (8) oropharyngeal width (OPhW; points G to H). Variable (9) is the segment vocal tract-oral (VT-O; points E to H). Reprinted with permission from [6]. Copyright 2009, Acoustical Society of America.

example, Vorperian and her colleagues compared the developmental trajectories of the nine variables they measured. They found the variables in the oral region, which expand along a horizontal plane and approximate mature adult size sooner than the variables in the pharyngeal region, which expand along a vertical plane [6]. The authors speculated that the different growth schedule of the structures in the oral/horizontal versus pharyngeal/vertical planes may reflect their diverse embryologic origins. Specifically, the embryologic structure type for the horizontal oral cavity is ectodermally derived, whereas the vertical pharynx region is endodermally derived. The difference in the growth speeds of vocal tract structures has been reported in other studies. For example, the hard palate reaches 80% of its mature adult length by the age of 18 months, whereas another structure, namely the pharynx, reaches 80% of its adult length by about the age of 6 years. The descent of the laryngeal and hyoid bone reached only 65% of the adult level at age 6 and continued descending considerably after that until they reached the mature adult position [5].

An anatomical observation of vocal tract development using MRI has also been reported by Fitch and Giedd [8]. They studied a relatively large data set (129 subjects/images) from 2 to 25 years old and compared the elongation of the lip, tongue blade, tongue dorsum, velum and pharynx in the prepubertal (2.8–8.1 yrs), pubertal (10.3–14.5 yrs), and postpubertal (14.7–17 yrs) periods. Although all of the five segments increased their length significantly across prepubertal and pubertal periods, only the velum and pharynx became larger during the pubertal and postpubertal periods. This fact coincides with the findings of Vorperian and her colleagues [5,6] that the structures of the vocal tract grow disproportionately rather than uniformly. In addition, there is a disproportionate increase in the pharyngeal length in both transitions. Fitch and Giedd compared the growth rate of each structure and reported that lip, tongue blade, tongue dorsum, and velum segments increase in size by an average of 12% between childhood and puberty, and only 5% between puberty and adulthood, whereas the pharynx length increases by 22% and 25%, respectively. Since this rapid elongation of the pharynx in the pubertal period is most pronounced in males, Fitch and Giedd pointed out that there are two phases during which large "descents of the larynx" take place: the first large descent occurs early in life, and continues gradually in accordance with the growth of the body, and a second large descent, which is restricted to males, occurs at puberty.

The pubertal descent of the larynx enlarges the vocal tract, especially the pharynx, of males. As a result, the formant frequencies of males after puberty become lower than those of females. The low formant frequencies as well as the low fundamental frequency (F_0) caused by longer vocal folds in males clearly distinguish the male voice from the female voice. However, it is well known that gender specific differences in acoustic features actually exist well before puberty during which there is a second rapid descent of the larynx [9–11]. Since there were no known anatomical differences between prepubertal boys and girls, the gender difference in the acoustic features in early stage of life was attributed to behavioral rather than anatomical factors. For example, boys may achieve a lower formant frequency by protruding their lips. By doing so, they can behaviorally imitate the long vocal tract that characterizes adult males.

However, in 2011 Vorperian and her colleagues [7] documented for the first time that there are statistically significant gender differences in the anatomy of the prepubertal vocal tract using 605 MRI and CT images, which were the same as those used in [6]. The result indicated that a gender difference in the vocal structure appears between approximately 3 and 7 years of age in a horizontal region ranging from a line tangential to lips to the posterior pharyngeal wall (vocal tract horizontal: VT-H). Interestingly, the difference subsequently disappears temporarily, and then reemerges at around 13–14 years. In contrast, the horizontal oropharyngeal width (OPhW) is slightly larger, albeit with no statistically significance, in boys approximately between 2 and 4 years of age, but the gender difference subsequently disappears permanently. These findings suggest that sexual dimorphism of the vocal tract structures is not consistently present (or absent) during the course of development. In addition, segments that expand toward a vertical plane display a significant gender difference later than those in the horizontal segments such as VT-H or OPhW, which are described above. For example, the nasopharyngeal length (NPhL) and the posterior cavity length (PCL) become longer in males than in females at approximately 8 years of age, and the vertical distance from the glottis to the palatal plane (VT-V) becomes longer in males after the age of 13. Therefore, the timeline for the emergence of gender dimorphism varies among the vocal tract segments. Another study by the same authors also indicated differences between the growth speeds in males and females. For example, the OPhW of males has grown to about 70% of its mature adult size by age 6, whereas that of females has grown to only about 50% of its mature adult size by the same age [6].

The anatomical development of the vocal tract can be summarized as follows, (1) the growth pattern and speed differ depending on each segment of the vocal tract, and horizontal segments mature faster than vertical segments, (2) a rapid larynx descent occurs twice during the course of development, and the latter pubertal descent occurs specifically in males, and (3) gender dimorphism can be observed in the horizontal segments of the vocal tract anatomy within a few years of birth. In addition, the growth patterns of males and females also differ for several segments.

3. ACOUSTIC CHANGE WITH THE DEVELOPMENT OF INFANT'S VOCAL TRACT

The shape of a vocal tract decides the resonant characteristics of its vocalizations. The peaks of the resonant frequencies are called formants, and the first two formants determine the vowel sounds. The formant frequencies correspond to the length and place of narrowing of the vocal tract. This means that the physical vocal tract development is reflected in the formant frequencies. For example, the formant frequencies are generally decreased by vocal tract enlargement and lip protrusion. Tongue positions also influence the formant frequencies. The first formant frequency (F_1) corresponds to the vertical height (high or low) of the tongue, while the second formant frequency (F_2) corresponds to the horizontal position (forward and backward) of the tongue [12]. On the other hand, the F_0 reflects the increases in the length and volume of a vocal cord [13].

There have been many previous studies that investigated acoustic development by analyzing the vocalizations of infants and children. One of the earliest investigations was undertaken by Eguchi and Hirsh [14]. They recorded two sentences read by each of 84 subjects ranging from 3 to 13 years old, and estimated F_1 and F_2 for the vowels /i/, /æ/, /u/, /ɛ/, /a/, and /ɔ/ in the sentences. The result showed that both F_1 and F_2 shifted downward with age for all the vowels except the F_1 of vowel /a/. The between-subject standard deviation of F_1 and F_2 also decreased with age. Moreover, they reported that the average F_0 is around 300 Hz at 3 years of age, and then decreases to around 250 Hz at 10 years old. After that, a female's F_0 decreases slightly from around 250 Hz to around 240 Hz whereas the male's F_0 decreases significantly from around 250 Hz to 220 Hz between 10 and 13 years old. Kent conducted a meta-analysis of relevant studies on formant frequencies including those of Eguchi and Hirsh [15], and suggested that the male's mean F_0 reaches the mature adult value (about 100 Hz) between 13 and 18 years of age. Lee *et al.* [16], who analyzed the speech of 436 children and 56 adults, also documented the decrement of formant frequencies and F_0s with age. In addition, they compared the vowel durations and spectral distances between children's and adults' tokens in the same vowel category, and revealed that the acoustic features converged to the canonical levels for adults at age 15. These acoustic changes are assumed to correspond to the enlargement of the vocal tract/folds as well as an improvement in articulation ability.

Lee and his colleagues also reported gender differences in acoustic features [16]. They found that the format frequencies decreased more rapidly in males than in females between 10 to 15 years of age and the gender difference in F_0 appeared between 11 and 15 years of age. Perry *et al.* also measured the first three formant frequencies of vowels at 4–, 8–, 12–, and 16 years of age [11]. They found that the F_3 values for boys were significantly lower than those for girls as young as 4 years of age. The same pattern was observed for the F_1 and F_2 values at 8 years and above. Furthermore, the gender difference in the three formant frequencies increased with age. Interestingly, although there is little difference between the F_0 values of boys and girls before 12 years of age, adult listeners could identify the speaker's gender even from the utterances of a 4 year old, and F_3 plays an important role for this identification.

As described in the previous section, several anatomical studies have reported that the vocal tract enlarges greatly in the first few years of life. This fact indicates that the acoustic features also change greatly during the same time period. Although there have been a considerable number of studies that reported a decrement of F_0 ([15,17] for review), scarcely any studies successfully observed a uniform decrement of formant frequencies during the period ([18–20], but see [21]). On the other hand, several studies have indicated an interesting developmental process in terms of vowel category formation. For example, Kent and Murray analyzed the resonant frequencies of vocalic utterances at 3, 6, and 9 months of age and demonstrated that the ranges of utterance in the F_1–F_2 plane expanded over time [18]. Kuhl and Meltzoff analyzed the vocalization of 3 to 5 month old infants when they were audiovisually presented with faces producing the vowels /a/, /i/ and /u/ [22], and found that the vowel categories became more separated in the vowel space with age. Gilbert and colleagues also analyzed utterances at 15 to 36 months of age and revealed that the formant values decreased or remained unchanged depending on the place of articulation of the vowels until 24 months of age [19]. In contrast, the formant frequencies decreased significantly regardless of the place of articulation from 24 months onwards.

Although the studies described above help us to understand the global developmental process of the formant frequency decrement and vowel space expansion, most of them employed cross-sectional data, which are not free from the negative effects of a wide inter-individual variability in early speech development [23,24]. Longitudinal data collection and analysis are required if we are to deal with the development process in detail. As regards longitudinal studies, Lieberman collected speech data produced by five infants aged from 4 to 60 months [25]. Although the sample size is very limited, his analysis of formant frequency plots over time revealed a gradual emergence of a well-developed vowel triangle with age. Buhr further provided a detailed analysis of a male infant in the sample from 4 to 16 months and revealed that the vowel space expands with age while there is no uniform decrement of the formant frequencies during the same period [26]. Bond and his colleagues also conducted a longitudinal analysis of a child's speech from 17 to 29 months and reported a similar developmental pattern [27].

More recently, Ishizuka and his colleagues analyzed a large-scale longitudinal database (the NTT Japanese infant speech database [28]) to provide detailed findings on the developmental process of speech production [29]. They estimated two lower resonant frequencies of more than 30,000 vowels in natural spontaneous speech for a male and a female child from 4 to 60 months of age. As result, they found that the F_1 values of /a/, /i/, /u/ and /e/ decreased with age until around 24 months except for vowel /u/ of the female infant. In contrast, the F_1 values for the same age interval in /o/ remained unchanged. Regarding F_2, the values for the front vowels, i.e., /e/ and /i/, increased or remained unchanged until 16 months of age, whereas the values for the central vowels /a/ and /u/, and the back vowel /o/ decreased until 20 months of age. Thus, the difference between the F_2 values of the front and central/back vowels increased with age. Furthermore, the vowel distributions on the F_1–F_2 plane indicated that the vowel space expanded until around the age of 18 months (Fig. 3), which coincides well with the findings of past studies. It is also indicated that, although there were inter-individual differences after 24 months of age, the developmental trajectories of the formant frequencies for the two infants were very similar until that age. The rapid develop-

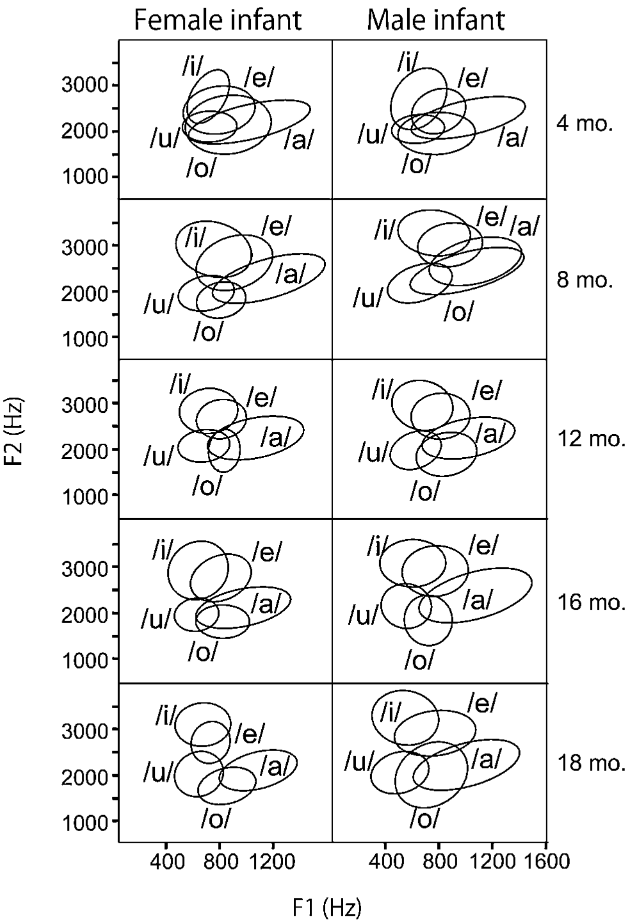

Fig. 3 Fifty percent probability ellipses for five Japanese vowels produced by two Japanese infants cast in the F_1–F_2 plane. The horizontal axis represents F_1 and the vertical axis represents F_2 in Hz. The vowel space expands as a function of age. The figure was recreated from [29].

ment in the first 2 years without any noteworthy inter-individual differences suggests that the process was caused mostly by certain universal developmental changes, e.g. the anatomical growth of the articulatory organs. Subsequently, factors such as manner of articulation and speaker gender begin to guide the inter-individual differences in acoustic features.

Modeling the speech production of children is also a promising approach for revealing the interaction between anatomical development and relevant acoustic change [30,31]. For example, Ménard and her colleagues modeled the vocal tracts of newborns and at 4, 10, 16 and 21 years of age by altering vocal-tract length of Maeda's model [32] which is an articulation-based speech synthesizer developed from the quantitative knowledge of the articulators. The acoustic space defined by the formant frequencies for each age group was obtained with randomly selected Maeda model parameters. Perceptual experiments on synthesized vowels with adult listeners revealed that all kinds of vowels in the French phoneme inventory could be perceived for the tokens generated even by a newborn vocal tract model. This fact indicates that the unique configuration of a newborn's vocal tract is assumed to be sufficient to produce any type of vowel.

However, the actual vowel inventory produced by infants is quite limited compared with the vowel inventory indicated by the simulation. This inconsistency may be explained in terms of immature articulation skill, e.g. the motor control of the jaws [31], and sophisticated articulation skills are probably needed to produce a full set of vowels. This idea agrees well with the fact that the vowel space expands steadily in the first couple of years. Nevertheless, further investigation is still required to prove the validity of the vocal tract model and to reveal the developmental process of articulation in detail.

We summarize the acoustic changes relevant to vocal tract development as follows; (1) A change in acoustic features is typically shown as a uniform decrement of F_0 and the formant frequencies at around 2 to 3 years of age and older. These acoustic features converge with the mature adult level at around 15 years of age. (2) An early change in the acoustic features until about 2 years of age appears as a separation of vowels or vowel space expansion rather than a uniform decrement of the formant frequencies. (3) As in the anatomical development, gender specific characteristics in the acoustic development are also documented during the first few years after birth.

4. CONCLUSION

The present paper reviewed the anatomical development and consequent changes in the acoustic features of utterances in children. There are several remaining questions that will stimulate further investigation.

As regards anatomical development, all the studies introduced in this paper measured vocal tract length using images of the midsagittal plane (a plane that divides the body vertically into left and right halves of equal proportion). However, the development of the vocal tract involves not only its length, but also its width and volume, and these must be taken into account. Studies using 3D images are needed to measure the vocal tract structure including these parameters, although we are not aware that any studies using 3D images of the developing vocal tract have yet been reported. Vocal folds are also an important anatomical structure, but their development is still not well known because of the difficulty of observing them in children. Since vocal folds affect not only F_0 but also the quality and emotion of the voice, further investigation is warranted.

In the latter half of this paper, we described the developmental change in formant frequencies. However, care must be taken that the changes in formant frequencies reflect more than anatomical development. In particular, the vowel categorization process observed in early childhood is presumably closely related to the maturation of articulation skills. Given the fact that infants can imitate the vowels they perceive as early as 5 months of age [22,33], the acquisition of articulatory movements is also expected to start in an early stage of development. A database of child articulation, just like those of adult articulation (e.g. [34]), will help us to untangle the developmental process. Recently, Zharkova and her colleagues have successfully observed the articulation of children using an ultrasonographic headset device, which will presumably contribute to the construction of a database of child articulation [35]. We also expect the child database to untangle the development of both vowel and consonantal articulation such as the tongue tip movement during /t/ production.

Another significant aspect is the neurological mechanism that underpins the articulation development. Guenther and his colleagues employed a constructivist approach to the problem. They proposed a computer simulation model called Directions Into Velocities of Articulators (DIVA), which learns a neural network representing the human brain structure by imitating ambient speech [36,37]. In this way, DIVA replicates the process of articulatory acquisition from the early babbling phase to later vowel/syllable production to provide unified explanations for the development of articulation.

Nonetheless, there are still problems to be solved if we are to develop a precise simulation model of articulatory acquisition, e.g. the inverse estimation of articulation from speech sound and an elaborate vocal tract model for children. We hope that the body of knowledge that has recently been accumulated at a rapid pace, as introduced in this paper, will help to untangle these remaining questions in the near future.

ACKNOWLEDGEMENTS

We thank Dr. Takemi Mochida (NTT CS Labs.) and Dr. Kentaro Ishizuka (NTT Communications) for their useful comments and suggestions.

REFERENCES

[1] V. E. Negus, *The Comparative Anatomy and Physiology of the Larynx* (Hafner, New York, 1949).
[2] U. G. Goldstein, "An articulatory model for the vocal tract of growing children," Doctor of Science thesis. MIT, Cambridge, Mass. (1980).
[3] D. E. Lieberman, R. C. McCarthy, K. M. Hiiemae and J. B. Palmer, "Ontogeny of postnatal hyoid and larynx descent in humans," *Arch. Oral Biol.*, **46**, 117–128 (2001).
[4] H. Takemoto and T. Kitamuara, "Overview of methods and techniques used in MRI-based speech production studies," *J. IEICE*, **94**, 585–590 (2011) (in Japanese).
[5] H. K. Vorperian, R. D. Kent, M. J. Lindstrom, C. M. Kalina, L. R. Gentry and V. S. Yandell, "Development of vocal tract length during early childhood: A magnetic resonance imaging study," *J. Acoust. Soc. Am.*, **117**, 338–350 (2005).
[6] H. K. Vorperian, S. Wang, M. K. Chung, E. M. Schimek, R. B. Durtschi, R. D. Kent, A. J. Ziegert and L. R. Gentry, "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," *J. Acoust. Soc. Am.*, **125**, 1666–1678 (2009).
[7] H. K. Vorperian, S. Wang, E. M. Schimek, R. D. Durtschi, A. J. Ziegert, L. R. Gentry and M. K. Chung, "Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study," *J. Speech Lang. Hear. Res.*, **54**, 995–1010 (2011).
[8] W. T. Fitch and J. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *J. Acoust. Soc. Am.*, **106**, 1511–1522 (1999).
[9] S. Bennet, "Vowel formant frequency characteristics of preadolescent males and females," *J. Acoust. Soc. Am.*, **67**, 231–238 (1981).
[10] J. Sachs, P. Lieberman and D. Erickson, "Anatomical and cultural determinants of male and female speech," in *Language Attitudes: Current Trends and Prospects*, R. W. Shuy and R. W. Fasold, Eds. (Georgetown University Press, Washington, DC, 1973), pp. 74–84.
[11] T. L. Perry, R. N. Ohde and D. H. Ashmead, "The acoustic bases for gender identification from children's voices," *J. Acoust. Soc. Am.*, **109**, 2988–2998 (2001).
[12] R. D. Kent and C. Read, *The Acoustic Analysis of Speech* (Singular Pub. Group, San Diego, 1992).
[13] M. Hirano, S. Kurita and T. Nakashima, "The structure of the vocal folds," in *Vocal Fold Physiology*, M. Hirano and K. N. Stevens, Eds. (University of Tokyo Press, Tokyo, 1981), pp. 33–41.
[14] S. Eguchi and I. J. Hirsh, "Development of speech sounds in children," *Acta Otolaryngol.*, **257**, 1–51 (1969).
[15] R. Kent, "Anatomical and neuromuscular maturation of the speech mechanism," *J. Speech Hear. Res.*, **19**, 421–447 (1976).
[16] S. Lee, A. Potamianos and S. Narayanan, "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *J. Acoust. Soc. Am.*, **105**, 1455–1468 (1999).
[17] S. Amano, T. Nakatani and T. Kondo, "Fundamental frequency of infants' and parents' utterances in longitudinal recordings," *J. Acoust. Soc. Am.*, **119**, 1636–1647 (2006).
[18] R. D. Kent and A. D. Murray, "Acoustic features of infant vocalic utterances at 3, 6, and 9 months," *J. Acoust. Soc. Am.*, **72**, 353–365 (1982).
[19] H. R. Gilbert, M. P. Robb and Y. Chen, "Formant frequency development: 15 to 36 months," *J. Voice*, **11**, 260–266 (1997).
[20] M. P. Robb, Y. Chen and H. Gilbert, "Developmental aspects of formant frequency and bandwidth in infants and toddlers," *Folia Phoniatr. Logop.*, **49**, 88–95 (1997).
[21] H. K. Vorperian and R. D. Kent, "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," *J. Speech Lang. Hear. Res.*, **50**, 1510–1545 (2007).
[22] P. K. Kuhl and A. N. Meltzoff, "Infant vocalizations in response to speech: Vocal imitation and developmental change," *J. Acoust. Soc. Am.*, **100**, 2425–2438 (1996).
[23] C. A. Ferguson, "Phonology as an individual access system: some data from language acquisition," in *Individual Differences in Language Ability and Language Behavior*, C. J. Fillmore, D. Kempler and W. S. Y. Wang, Eds. (Academic Press, New York, 1979), pp. 189–201.
[24] C. A. Ferguson and C. Farwell, "Words and sounds in early language acquisition," *Language*, **51**, 419–439 (1975).
[25] P. Lieberman, "On the development of vowel production in young children," in *Child Phonology*, Vol. 1 Production, G. H. Yeni-Komshian, J. F. Kavanagh and C. A. Ferguson, Eds. (Academic Press, New York, 1980), pp. 113–142.
[26] R. D. Buhr, "The emergence of vowels in an infant," *J. Speech Hear. Res.*, **23**, 73–94 (1980).
[27] Z. S. Bond, L. Petrosino and C. R. Dean, "The emergence of vowels: 17 to 26 months," *J. Phonet.*, **10**, 417–422 (1982).
[28] S. Amano, T. Kondo and K. Kato, "Development of NTT infant speech database," (in Japanese) *IEICE Tech. Report*, TL, **108**(50), pp. 29–34 (2008).
[29] K. Ishizuka, R. Mugitani, H. Kato and S. Amano, "Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infants," *J. Acoust. Soc. Am.*, **121**, 2272–2282 (2007).
[30] L. Ménard, J. Schwartz and L. J. Boë, "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood," *J. Speech Lang. Hear. Res.*, **47**, 1059–1080 (2004).
[31] J. E. Serkhane, J. L. Schwartz, L. J. Boë, B. L. Davis and C. L. Matyear, "Infants' vocalizations analyzed with an articulatory model: A preliminary report," *J. Phonet.*, **35**, 321–340 (2007).
[32] S. Maeda, "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, W. Hardcastle and A. Marchal, Eds. (Kluwer Academic, Dordrecht, the Netherlands, 1990), pp. 131–149.
[33] M. Patterson and J. F. Werker, "Matching phonetic information in face and voice is robust in 4.5-month-old infants," *Inf. Behav. Dev.*, **22**, 237–247 (1999).
[34] M. Hashi, "The X-ray microbeam speech production database in Japanese," (in Japanese) *J. Phonet. Soc. Jpn.*, **4**(2), pp. 31–35 (2000).
[35] N. Zharkova, N. Hewlett and W. J. Hardcastle, "Coarticulation as an indicator of speech motor control development in children: An ultrasound study," *Mot. Control*, **15**, 118–140 (2011).
[36] F. H. Guenther, "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," *Psychol. Rev.*, **102**, 594–621 (1995).
[37] D. E. Callan, R. D. Kent, F. H. Guenther and H. K. Vorperian, "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," *J. Speech Lang. Hear. Res.*, **43**, 721–736 (2000).