Infant Behavior & Development 46 (2017) 178–193

Contents lists available at ScienceDirect

Infant Behavior and Development

Full length article

Acquisition of vowel articulation in childhood investigated by acoustic-to-articulatory inversion

Hiroki Oohashi*, Hama Watanabe, Gentaro Taga

Graduate School of Education, The University of Tokyo, Bunkyo-ku 7-3-1, 113-0033 Tokyo, Japan

ARTICLE INFO

Article history:
Received 25 October 2015
Received in revised form
28 November 2016
Accepted 31 January 2017
Available online 20 February 2017

Keywords:
Development
Acoustic-to-articulatory inversion
Articulation
Speech

ABSTRACT

While the acoustical features of speech sounds in children have been extensively studied, limited information is available as to their articulation during speech production. Instead of directly measuring articulatory movements, this study used an acoustic-to-articulatory inversion model with scalable vocal tract size to estimate developmental changes in articulatory state during vowel production. Using a pseudo-inverse Jacobian matrix of a model mapping seven articulatory parameters to acoustic ones, the formant frequencies of each vowel produced by three Japanese children over time at ages between 6 and 60 months were transformed into articulatory parameters. We conducted the discriminant analysis to reveal differences in articulatory states for production of each vowel. The analysis suggested that development of vowel production went through gradual functionalization of articulatory parameters. At 6–9 months, the coordination of position of tongue body and lip aperture forms three vowels: front, back, and central. At 10–17 months, recruitments of jaw and tongue apex enable differentiation of these three vowels into five. At 18 months and older, recruitment of tongue shape produces more distinct vowels specific to Japanese. These results suggest that the jaw and tongue apex contributed to speech production by young children regardless of kinds of vowel. Moreover, initial articulatory states for each vowel could be distinguished by the manner of coordination between lip and tongue, and these initial states are differentiated and refined into articulations adjusted to the native language over the course of development.

1. Introduction

The speech sounds are generated by complex motor coordination among the articulatory organs. While the developmental process of speech production has previously been depicted mainly on the basis of evidence derived from acoustical phenomena and their consequences—such as spectral envelope, fundamental frequencies (Amano, Nakatani, & Kondo, 2006; Ishizuka, Mugitani, Kato, & Amano, 2007; Kent & Murray, 1982; Vorperian & Kent, 2007) and phonetic transcriptions (Ingram, 1974; MacNeilage, 2000; MacNeilage & Davis, 2000; Oller, 2000; Stoel-Gammon & Cooper, 1984) —the development of the articulatory system by which these acoustics are produced still remains an open question because of limitations on the measurement of the articulatory system, especially that of tongue movements. In the present paper, we investigated longitudinal changes in children's articulation by estimating the parameters of an articulatory model on the basis of the acoustical features of speech sounds.

---

* Corresponding author.
E-mail address: hiro084@gmail.com (H. Oohashi).

http://dx.doi.org/10.1016/j.infbeh.2017.01.007
0163-6383/

The development of speech production during the first year of life has been characterized as following a particular course (Kuhl, 2004; Oller, 2000; Stoel-Gammon & Cooper, 1984). Infants are born able to produce spontaneous sounds, such as sneezing and crying. Infants then produce cooing, that is, quasivocalic sounds similar to vowels. Subsequently, coos expand into clear vowel sounds characterized by full resonance and wide variety. At an early stage of babbling, a large portion of sounds produced by infants can be heard as repetitions of the same consonant–vowel (CV) units such as "papapa" and "mamama." After that stage, infants combine different consonant- and vowel-like sounds to produce variegated sequences. Finally, beginning around the end of the first year of life, infants produce meaningful speech.

Acoustical studies show that as children grow up, their vowel clusters become more distinct, and the fundamental frequency and spectral peaks (formant frequencies) of their utterances become lower (Amano et al., 2006; Ishizuka et al., 2007; Kent & Murray, 1982; Vorperian & Kent, 2007). Moreover, analyses of phonetic transcriptions show a modification process at work in infants' vocalizations (MacNeilage, 2000; MacNeilage & Davis, 2000). At the babbling stage, infants prefer to repeat three predominant CV sequences, that is, labial–central, coronal–front, and dorsal–back CV patterns. With development, children begin to chain variegative CVs, with a fronting tendency in which the first consonant in words has a more anterior place of articulation than the second one (Ingram, 1974). These phenomena are crosslinguistically observed (Amano et al., 2006; Ishizuka et al., 2007; Kent & Murray, 1982; MacNeilage, 2000; MacNeilage & Davis, 2000; Vorperian & Kent, 2007).

These changes are likely to be caused mainly by the development of vocal tract anatomy, respiration, and motor controls of articulators. In order to investigate the anatomical structure of the articulatory system and its dynamics during speech production, previous studies have adopted a variety of methods, such as radiographic imaging (Chiba & Kajiyama, 1942; Fant, 1960; Kiritani, 1986), electromagnetic articulography and electropalatography (Byrd & Tan, 1996; Hixon, 1971), magnetic resonance imaging (Fitch & Giedd, 1999; Masaki et al., 1999; Vorperian, Kent, Gentry, & Yandell, 1999; Vorperian et al., 2005), ultrasound (Geddes, Kent, Mitoulas, & Hartmann, 2008; Zharkova, Hewlett, & Hardcastle, 2011), and motion-capture systems (Green, Moore, Higashikawa, & Steeve, 2000; Green, Moore, & Reilly, 2002; Goffman & Smith, 1999; Nip, Green, & Marx, 2009). With regard to anatomy, previous studies reveal that children's vocal tracts, especially during the first year of life, are shaped differently from those of adults (Fitch & Giedd, 1999; Goldstein, 1980; Sasaki, Levine, Laitman, & Crelin, 1977; Vorperian et al., 1999, 2005). Infants' vocal tracts are not only smaller than adults', but they have a relatively lager oral cavity than pharyngeal one, a flat tongue, and a more gradually sloping pharyngeal tract. These properties of the infant vocal tract should raise formant frequencies and lead to less clear vowel clusters. In addition, the limited range of tongue movement prevents complex consonantal articulations. While these anatomical changes in vocal tract are certainly responsible for the changes in the filter properties of speech sounds, their phonation is conversely affected mostly by the development of respiration (Boliek, Hixon, Watson, & Morgan, 1996; Reilly & Moore, 2009). For instance, decrease in the compliance of the chest wall results in more rapid modulation of respiratory muscle movements.

As for the development of motor control of articulators, transcription analysis suggests that infants have relatively independent control over their jaw and that ability to carry out tongue movements depends largely on jaw control (MacNeilage, 2000; MacNeilage & Davis, 2000). On the basis of these findings, it has been convincingly argued that mandibular oscillations have a crucial role in the early development of articulation. One study using motion capture partly supports this idea by reporting that jaw movements mature earlier than lip ones (Green et al., 2002; Nip et al., 2009). Another study, using electromagnetic articulography and acoustical analysis, reports that fronting tendencies that are predominant in both adults and children are caused by coordination among articulators (Rochet-Capellan & Schwartz, 2007).

Thus, as described above, the acoustical analysis and empirical measurement of the articulatory system reveals much about the development of speech production. Taking into consideration that vowel production accounts for a large portion of speech by young children, tongue movements would play a crucial role in development of speech production. However, many aspects of the development of articulation, especially tongue movements during speech production until the second year of life, still remain an open question. This is because of limitations to the empirical measurement of articulatory movements in young children.

Another approach to investigate articulatory movements is to estimate articulatory states from acoustical features; this is called acoustic-to-articulatory inversion (Atal, Chang, Mathews, & Tukey, 1978; Hiroya & Honda, 2004; Ménard, Schwartz, & Boë, 2004; Ouni & Laprie, 2005; Shirai, 1993; Toda, Black, & Tokuda, 2008; Uchida, Saito, Minematsu, & Hirose, 2015; Uria, Renals, & Richmond, 2011; Wakita, 1973). This technique relies on a mapping function from acoustical to articulatory space. Previous studies have proposes several such mapping functions (Atal et al., 1978; Hiroya & Honda, 2004; Ouni & Laprie, 2005; Shirai, 1993; Wakita, 1973) and, on their basis, articulatory models (Maeda, 1990; Mermelstein, 1973; Story, 2009). When it comes to applying this technique to sounds produced by infants, however, some problems arise. First, because of anatomical differences between infants' vocal tracts and those of adults, the articulatory model used must be scalable to the child's vocal tract size. Second, we cannot calculate a mapping function from acoustical to articulatory features, since it is impossible to pair acoustical features with empirically obtained articulatory features in this case. Third, although the model should approximate the vocal tract shape, it is desirable to have a smaller number of parameters.

Taking into consideration the need for scalability of the vocal tract and parameters to specify articulatory states, we adopted Maeda's model (Maeda, 1990; Ménard et al., 2004; Serkhane, Schwartz, Boë, Davis, & Matyear, 2007). This model was proposed to approximate midsagittal slices of the vocal tract during adult' vowel productions (Maeda, 1990). Subsequent studies (Ménard et al., 2004; Serkhane et al., 2007) propose two scaling factors to incorporate growth data (Goldstein, 1980) into the model and apply it to non-adult-sized vocal tracts. A previous study (Serkhane et al., 2007) compares simulated