Journal of Phonetics 71 (2018) 16–34

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Cross-linguistic differences in the size of the infant vowel space

Abdulsalam Alhaidary [a,*], Susan Rvachew [b]

[a] Department of Rehabilitation Sciences, College of Applied Medical Sciences, King Saud University, P.O. Box 10219, Riyadh 11433, Saudi Arabia
[b] School of Communications Sciences and Disorders, Faculty of Medicine, McGill University, 2001 McGill College Avenue, 8th Floor, Montréal, Québec H3A 1G1, Canada

ARTICLE INFO

Article history:
Received 27 August 2017
Received in revised form 17 May 2018
Accepted 20 July 2018

Keywords:
Speech development
Ambient language
Vowel space
Babble
Babbling drift

ABSTRACT

This study examined the effects of linguistic environmental input on babbling in cross-linguistic investigations of vowel space. Speech samples were collected from 10- to 18-month-old infants learning Arabic ($N = 31$). First (F1) and second (F2) formant frequencies were identified in the selected vowels and used to calculate the compact-diffuse (F2 – F1) and grave-acute ([F2 + F1]/2) values for each vowel and the size of the vowel space was calculated for each infant's vowel space. These vowel space statistics were compared to similar data derived from vowels produced by English-learning infants ($N = 20$) and French-learning infants ($N = 23$) as previously described in Rvachew, Mattock, Polka, and Menard (2006). It was found that Arabic infants appeared to achieve a larger vowel space at a younger age compared to the English and French infants, which we attribute to the benefit of a less crowded vowel space in Arabic input compared to English and French input. Expansion of the vowel space toward the diffuse and grave corners was common to all three language groups, but the developmental trajectories for the mean F1 and mean F2 varied with language input. These findings suggest that the development of infant babbling is influenced by a complex interaction of endogenous and exogenous processes, which include the biological development of the vocal tract and language input from the ambient environment.

1. Introduction

Acoustic studies of infant speech production have identified the expansion of the vowel space along its F1 and F2 dimensions as a fundamental developmental accomplishment during the first 18 months of life (Vorperian & Kent, 2007). Fully resonant vowels emerge early in an infant's vocal repertoire and increase rapidly in frequency after 4 months of age (Oller, 2000). Initially, the vowels produced by an infant tend to be low, front, and central vowels (e.g., [ɛ], [æ], [ə]), even though production of the full vowel repertoire is not precluded by infant vocal tract anatomy (Buhr, 1980a, 1980b; Ménard, Schwartz, & Boë, 2002). With increasing age, vowels from the periphery of the vowel space are produced more often, as judged by adult listeners (Rvachew, Alhaidary, Mattock, & Polka, 2008). The acoustic correlate of the addition of these peripheral vowels is a gradual increase in the range of F1 values that occurs during the first year (Kent & Murray, 1982), followed by an increase in the range of F2 values that occurs primarily during the second year (Rvachew, Slawinski, Williams, & Green, 1996). Together, the expanded range of these formant values corresponds to an increase in the size of the vowel quadrilateral.

Most often this expansion is attributed to the maturation of speech motor control. For example, Buhr (1980a, 1980b) speculated that the later appearance of /u/ and other rounded vowels reflected the need for coordinated articulation of the lips, jaw, and tongue. Generally, it is accepted that early infant vocalizations are characterized by undifferentiated movements of these articulators, with vocal output dominated by cyclic movements of the jaw (termed *frame dominance* from the perspective of the frame-content theory of speech development, c.f., MacNeilage, 1998). As reviewed by Rvachew and Brosseau-Lapré (2018), electromyographic, kinematic, and acoustic studies have confirmed that jaw stability for speech emerges from an initial state of poorly controlled muscle activation patterns; subsequently, lip and tongue movements are coordinated with the dominant jaw movement pattern through the processes of differentiation and integration, which result in the production of increasingly mature speech syllables (e.g., Green, Moore, Higashikawa, & Steeve, 2000).

Although the immaturity of speech motor control may initially limit the vowel repertoire and play a role in the late

---

* Corresponding author.
E-mail address: aalhaidary@ksu.edu.sa (A. Alhaidary).

https://doi.org/10.1016/j.wocn.2018.07.003
0095-4470/© 2018 Elsevier Ltd. All rights reserved.

appearance of peripheral vowels, it is clear that the maturation of the articulatory function by itself cannot explain the expansion of the vowel space during early speech development. Across multiple domains of motor development, empirical research has highlighted the importance of practice with feedback for the acquisition of motor skill (for further discussion and data on the speech domain, see Walsh, Smith, & Weber-Fox, 2006). Infant speech development is highly sensitive to alterations in auditory feedback. On the one hand, the emergence of canonical (i.e., speech-like) babble is inevitably delayed in the case of severe sensory neural hearing impairment (Bass-Ringdahl, 2010; Eilers & Oller, 1994), whereas on the other hand, precanonical vocalizations appear to be unaffected by the absence of auditory feedback. These findings have led Koopmans-van Beinum, Clement, and van den Dikkenberg-Pot (2001) to conclude that auditory input is essential for learning to coordinate multiple articulators. A specific effect of auditory deprivation, even in the case of transient hearing loss in infancy, is a persistently restricted vowel space (Kent, Osberger, Netsell, & Hustedde, 1987; Rvachew et al., 1996), which contrasts with the gradual enlargement of the vowel space in infants with normal hearing.

Various learning mechanisms have been proposed to explain the essential role of auditory feedback in the emergence of speech-like vocalizations during the first year of life. One possibility is that canonical babbling is associated with supervised learning in which a target output is specified; during practice, feedback generates error signals that are used to improve performance (Wolpert, Ghahramani, & Flanagan, 2001). Traditionally, supervised learning invokes the notion of an external teacher, and at least one study has suggested imitation of point vowels occurs by infants as young as 6 months of age (Kuhl & Meltzoff, 1996). Self-supervised learning also is possible, especially given the early stabilization of native-language vowel categories in perceptual learning (Kuhl et al., 2008). Moulin-Frier, Nguyen, and Oudeyer (2014) have described a model in which early learning is motivated intrinsically and focused on self-generated auditory targets; a developmental transition to imitation learning occurs later in development after the achievement of the basic principles of speech production. In contrast, Howard and Messum (2011) have suggested that reinforcement learning is the primary mechanism, with speech learning driven by adult mimicry of infant vocalizations that capture adult attention when they approximate phonetic categories in the adult language system (see also Rasilo, Räsänen, & Laine, 2013). These differing learning mechanisms are not mutually exclusive, and it is likely that all three play a role in early speech development.

Regardless of the learning mechanism that may drive an infant's vocal output at any given time, the relative salience of specific phonetic categories may interact with articulatory constraints to explain the specific vocalizations that are produced. In supervised learning, the infant is essentially responsible for selecting targets for attention. In reinforcement learning, the adult selects infant vocalizations that are speech-like, but the infant must recognize the correspondence between the self-produced utterance and the adult imitation, and ultimately, both speaking partners must be attracted by the salience of the utterances in question. Polka and Bohn (2011) have proposed the Natural Referent Vowel framework to account for universal patterns of asymmetry in vowel perception whereby infants and adults in predictable circumstances are more able to perceive a vowel change from a more central to a more peripheral vowel than from a more peripheral to a more central vowel. These authors have speculated that a perceptual bias toward these referent vowels enables an infant to establish an early minimal vowel repertoire at the periphery of the vowel space. These referents then serve as anchors from which the remaining vowel categories in the ambient language can be abstracted. Schwartz, Abry, Boe, Menard, and Vallee (2005) have presented evidence that the corner vowels are perceptually salient because of the focalization of acoustic energy when the formants converge (F1 – F2 for /u/, F2 – F3 for /a/, and F3 – F4 for /i/). Their Dispersion-Focalization Theory predicts optimum vowel sets given the size of the vowel system on the basis of perceptual factors that include the dispersion of vowels across the system and the focalization of energy within individual vowels. Their modeling research suggests that F1 is weighted more than F2, a parameter that contributes to a preponderance of symmetrical vowel systems with a small number of peripheral vowels (Schwartz, Boe, Vallee, & Abry, 1997).

Cross-linguistic research has revealed complex interactions between universal perceptual biases and the details of speech input in the development of native language perceptual categories during infancy. Research has shown that specific speech input can induce, maintain, or enhance infant perception of phonetic contrasts in ambient language (for a review, see Polka, Rvachew, & Mattock, 2007), with language-specific perception of vowel contrasts achieved early in comparison to consonant contrasts. Research also has shown that infant speech perception performance is influenced by the salience of the acoustic cues involved, as well as the infant's specific experience with the range of acoustic cues associated with a given phonetic category (Mattock, Polka, Rvachew, & Krehm, 2010). In contrast to the rich body of cross-linguistic research that exists with respect to the perceptual domain, cross-linguistic studies of infant speech production are less common but necessary to understanding the interaction between universal constraints on speech production and environmental input as explanations for individual differences in speech production learning.

In two prior studies, we described the development of the vowel space by infants who were learning Canadian English or Canadian French. In a cross-sectional study, Rvachew, Mattock, Polka, and Menard (2006) recorded vowels produced by 43 infants aged between 300 and 547 days. Each infant's vowel space was described in terms of raw acoustic parameters (mean and standard deviation of F1 and F2, in mels), which were used in turn to derive the features diffuse [max(F2 – F1), grave [min(F1 + F2)/2], acute [max(F1 + F2)/2], and compact [min(F2 – F1)]. These features, proposed by Jakobson, Fant, and Halle (1963), have been used to define the corners of the vowel space and to relate perceived acoustic features to produced vowels, with /i/ being in the *diffuse* corner, /u/ in the *grave* corner, /æ/ in the *acute* corner, and /a/ in the *compact* corner of the space. For example, Kuhl et al. (1997) have used this procedure to demonstrate that the vowel space is expanded in infant-directed speech relative to adult-directed speech in three language groups. In our application, the