Evolutional View of Preverbal Vocalizations in Human Infants

Miki TAKAHASI*, **

SUMMARY: Preverbal vocalizations in human infants drastically change during their first year. Previous studies have classified preverbal vocalizations into several stages based on the acoustical characteristic and the age. Vocal learning is the acquisition of new vocal patterns based on social and environmental experiences, and some animals are reported to exhibit vocal learning. In this paper, we view language as sequential vocal patterns and will discuss the developmental change in preverbal vocalizations in human infants from an evolutionary point of view.

Key words: vocal learning, development, call, song, evolution

1. Introduction

How do human infants acquire language? In the first few months, infants mostly produce crying sounds. After several months, they produce variable sounds, some of which are like language. At approximately 12 months, the first words are formed. The preverbal vocalizations in human infants drastically change during their first year. Oller (1980) and Stark (1980) classified preverbal vocalizations into four stages based on the acoustical characteristics and the age at production. The first stage is “phonation,” which is followed by “cooing” or “going,” then expansion or vocal play, and finally, babbling (canonical and variegated babbling) appears. In particular, babbling is considered the starting point of language acquisition because it contains syllables with consonants and vowels, and these syllables are repeated in babblings. However, the relationship between pre-babbling vocalizations and language acquisition remains unclear. Why do infant vocalizations change during development? What mechanisms underlie the developmental change of vocalizations into language?

Human language is a specific vocalization system that differs from other animal vocalizations. Therefore, it is difficult to directly compare the vocal development of human infants with that of other animals. In this paper, we consider language as sequential vocal patterns. Vocal learning is the acquisition of new vocal patterns based on social and environmental experiences (Janik and Slater 2000), and some animals are reported to exhibit vocal learning. The five groups of mammals that are vocal learners are whales and dolphins, seals, bats, elephants, and humans. In birds, three groups, songbirds, hummingbirds, and parrots, are known as vocal learners. There are two types of learned vocalizations. One is the song, which has a hierarchical structure of several elements that are acoustically complex sounds. The functions of songs are mainly limited to a reproductive context: courtship and the defense of territories. Another is the calls, which are acoustically simple and monosyllabic. Often, one syllable is repeated and calls function as a social tool. There are several types of calls, and each call has a specific use. For example, flight calls are used when flying off, alarm calls to announce the presence of predators to social members, and contact calls for informing of the existing site towards companies (Marler 2004). Contact calls are learned in several animal species. Dolphins only learn calls, but parrots learn both songs and calls. There are also strain differences with learned vocalizations.

Song learning in songbirds is comparable to human language acquisition, and there are several similar characteristics between them (Doupe and Kuhl 1999). Shared characteristics between the two exist in the following areas: the critical period to learn vocal sequences with a hierarchical structure, the developmental change, and social interaction for facilitating learning. We will review song learning and call learning, with a focus on the social structure and the context in which

* RIKEN Brain Science Institute (RIKEN脳科学総合研究センター)
** Research Fellow of Japan Society for the Promotion of Science（日本学術振興会特別研究員）
2. Song Learning

2.1 Song Learners

Acoustical similarity of song elements and similarity of transition patterns are required for song learning. Whales (only baleen, not toothed whales) and three strains of birds are known as song learners. Although the developmental change involved in song learning in whales remains unclear, longitudinal recording studies suggest that whales learn songs. Humpback whales migrate to warmer oceans before the breeding season and then return north where they remain until the next breeding season. Recordings over many years have revealed that male humpback whales sing in the breeding area (Payne and McVay 1971). They share song elements with members of their migration groups, and songs slightly change with breeding season (Noad et al. 2000).

Furthermore, it has been found that song changes are transmitted to other groups in distant areas (Garland et al. 2011). These studies strongly suggest that whales can learn songs. Study of whale songs is difficult because of their underwater habitat and long-distance, year-round migration patterns. There are remaining questions as to how long whales maintain the plasticity for song learning, and what mechanisms mediate song learning. Songbirds are very popular with people and the total number of songbird species accounts for more than half of all bird species. Therefore, numerous studies have investigated their song learning and its function. In this section, we focus on the song learning of the Bengalese finches, one of the songbird species.

2.2 Sensory and Sensory-motor Learning

Male Bengalese finches sing to females during courtship, a behavior that they learn within a few months after hatching. There are two stages in song learning. In the sensory learning period, juveniles memorize tutor songs by auditory inputs. Their songs gradually become similar to the tutor songs. When juveniles do not hear tutor songs due to being socially isolated or deafened, their songs become abnormal. In Bengalese finches, the auditory experience is required during the first 35–70 days after hatching (Clayton 1987). In the sensory-motor learning stage, the birds begin to sing. They modify their songs by auditory feedback to match their tutor songs. The early songs are immature and are called “subsongs.” In “plastic songs,” song elements become clear, but transition patterns are still unstable. Then, transition patterns become stable, and at last songs are crystalized (“crystalized songs” or “full songs”).

Longitudinal recordings of juvenile songs have revealed the developmental trajectory of song elements and transition patterns. In the beginning, immature song elements vary and are diverse in the acoustic space. Gradually, elements become acoustically clear and stable (Tchernichovski et al. 2001). There are two types of acoustic developments (Liu et al. 2004). One is the bottom-up style where the elements gradually develop identical acoustic features. The other type is the top-down style where immature sounds are differentiated into each song element. Once each song element is stable, researchers can categorize them based on the acoustic features. Lipkind et al. (2013) showed that paired song elements are fundamental for the development of transition patterns. Zebra finches, another songbird species, and Bengalese finches acquire song transition patterns by accumulating paired elements in a step-wise manner.

 Songs require volitional control of respiration for sequential vocalizations. Temporal parameters of songs are related to respiration control and can change throughout the bird’s development. The note duration and the note-to-note interval, which is the silent interval between two song elements, vary in the early phase. The timing of expiration vocalization is not synchronized with the onset of sound in the early subsong phase (Veit et al. 2011). The note duration stabilizes, and its distribution converges earlier than that of the note-to-note intervals. In the development of songs, acoustic features and note duration become stable first, followed by transition patterns and note-to-note intervals. Taken together, control of respiration for sequential vocalization and transition pattern might be related (Sasahara et al. 2015).

2.3 Sexual Selection and Song Learning

Songs are performed under the reproductive context and female birds have a specific preference for complex songs. Therefore, songs could evolve by sexual selection. In Bengalese finches, son’s song share only 80% of their song elements with their father’s songs (Takahasi and Okanoya 2010). Under the multi-tutor condition, most of juveniles select two males as tutors and copy some parts of their songs which have a high transition probability (Takahasi et al. 2010). However, the copied parts of the tutor song as well...
as the individual birds that are selected as tutors vary among juveniles. These individual differences might be related to female preferences. Female birds tend to prefer complex songs. Nowicki et al. (2002) showed that early nutritional conditions affect the accuracy of song learning, the ability to form complex songs, and the development of the brain region used for song control. In most cases, song learning occurs only once early in the lifetime of the bird. Therefore, females can assess how well the male has grown up just by listening to his song.

3. Call Learning

3.1 Bottlenose Dolphin Case Study

Although song learning is reported in only birds and whales, call learning is reported in most of the vocal learners. Bottlenose dolphins have a diverse vocal repertoire: clicks for echolocation, whistles, buzzes, quacks, and pops. The whistle is narrow-banded with frequency modulations and is a learned call. Dolphin calves start to produce burst pulses within 24 to 48 hours after birth. They also produce whistles approximately 5 days after birth (Killebrew et al. 2001). Whistles are not stable, and their pattern of frequency modulation varies during their first several months (Caldwell and Caldwell 1979, Tyack and Sayigh 1997) and can change drastically during their first year. Whistles with unstable frequency modulation are observed in calves across the different social groups. Over several months, dolphin calves produce whistles with identical frequency modulate patterns. Furthermore, they learn to produce shared whistles in their social group until they are one year old. Individual-specific whistles are thought to be used for individual identity and, therefore, are called signature whistles (Caldwell and Caldwell 1979). Dolphins still produce variable whistles after acquiring the signature whistle. Variable whistles involve the signature whistles of other individuals. It was discovered that members of alliances learn each other’s signature whistles (King et al. 2014). The development and learning of whistles are affected by social factors; therefore, the timing of learning differs between individuals. Adult dolphins also can mimic artificial whistles and learn the association between objects and sounds (Richards et al. 1984). This suggests that dolphins maintain a high plasticity of vocalization even as adults.

3.2 Bat Call Learning

Bats have several vocal repertoire in addition to a pulse for echolocation. Pup sac-winged bats produce isolation calls when they separate from their mothers. These isolation calls are learned by young bats. Bat pups have limited mobility for approximately 2 to 3 weeks after birth. During this period, the isolation calls differ acoustically between individuals, and mothers respond to the isolation calls of their offspring (Knörnschild and von Helversen 2008). The individual differences in isolation calls make it easy to find their offspring. While juveniles can fly by themselves 4 to 8 weeks after birth, they are still nursing and producing isolation calls. Also during this period, individual differences in isolation calls are reduced and the calls become increasingly similar to those shared with colony members (Knörnschild et al. 2012). In the same period, juveniles produce vocal sequences including immature and adult-like calls. These vocal sequences are similar to the songs that adult males produce during courtship and territorial defense. These vocal sequences might facilitate motor learning (Knörnschild et al. 2006), but more research is needed to confirm this point.

Other bats also learn calls, and in some species, call learning has been observed in adults. In spear-nose bats, non-related adult females gather for foraging. They communicate with a screech call when they gather, and these screech calls are shared between colony members (Boughman 1998).

3.3 Why Do They Learn Calls?

Call learning is useful for discriminating between colony members and outsiders. Dolphins have a fission-fusion group. However, they each form an alliance with a specific individual (Connor et al. 2001), and the members learn each other’s signature whistles. Sac-winged bats create harems where a dominant male mates with several females. The dominant male sometimes attacks subordinate males. When the subordinate males produce isolation calls shared in the colony, the dominant male stops attacking. The production of the shared isolation call indicates that the individual is a descendant of the dominant male and results in a confirmation of the paternity (Knörnschild et al. 2012). Parrots learn calls as well as warble songs for courtship. Calls are shared between mates or a social group (Sewall et al. 2016). Pinnipeds (seals, sea lions, and walruses) and elephants also learn calls (Reichmuth and Case 2014, Stoeger and Manger 2014). While walruses and elephants mimic the vocalizations of their specific caretakers, it is not clear how they use learned calls in the wild. These animals live in stable colonies for a long time, similar to dolphins and bats. This suggests that call learning might have evolved
Third, the gene expression pattern in the premotor area deteriorated following damage to the premotor area. Second, their song becomes intermediate neural areas that are homologous to the forebrain of songbirds. First, they have neural connections between neural areas that are homologous to the forebrain of songbirds. Recent studies have suggested that the transition pattern might be influenced by prenatal and postnatal conditions (Zhang and Ghasanfar 2016).

They are considered non-vocal learners, similar to pigeons, because their songs are not affected by deafening procedures and social isolation and they do not have well-formed nuclei in forebrain nuclei, which are related to the production of songs in songbirds. Recently, it was found that the phoebe (a suboscine bird) has a long-lasting sensory-motor period and that their songs gradually change during development. Furthermore, phoebes showed several similar neurological phenomena to songbirds. First, they have neural connections between neural areas that are homologous to the forebrain pathway in songbirds. Second, their song becomes deteriorated following damage to the premotor area. Third, the gene expression pattern in the premotor area is similar to that of songbirds. These results suggest that suboscine birds have a rudimentary sensor-motor learning system that is shared by songbirds before splitting from common ancestor (Liu et al. 2013).

5. Vocal Development in Human Infants

This section will review the development of preverbal vocalization in human infants. Preverbal vocalizations of infants were divided into four stages by Oller (1980) and Stark (1980). The first stage is phonation, which consists of all non-reflective vocalizations except crying. Vocalizations in this stage are accompanied by vibration of the vocal folds, but there is an insufficient resonance of the vocal tract. These vocalizations are observed in 0–2 months old, but are infrequent. During this period, infants frequently produce crying sounds. Other animals also have vocalizations for eliciting parental behavior comparable with crying in human infants. Most of them are sequential vocalizations; for example, begging calls in birds, isolation calls in bats, and babbling in marmosets. Mice (Grimsley et al. 2011) and giant otters (Mumm and Knönschild 2014) also produce sequential vocalizations to prompt parental behavior. It is potentially easy for infants to change signal duration or amplitude depending on the condition or demand in sequential vocalizations, and these signal changes alert caretakers (as receivers) to the change in the condition of the infant. Babbling in common marmosets is closely related to the respiration pattern (Zhang and Ghazanfar 2016) and begging calls are affected by neural pathways for song control (Liu et al. 2009). Sequential vocalizations also facilitate the development of respiration control for vocalizations. In human infants, crying may lead to volitional control over respiration and helps to form the foundation for the subsequent sensory-motor learning phase.

5.1 Function of Cooing

Cooing is frequently observed in two-month-old infants. This vocalization is not produced sequentially, while vocalizations in the previous and the following stages are sequential sounds. However, mothers or caretakers vocally respond to cooing and these vocal interactions occur repeatedly. Gratier et al. (2015) reported that two- to five-month-old infants show vocal turn-taking with their mothers. Infants and their mothers interact with vocal imitation and matching (Kuhl and Meltzoff 1996, Papousek and Papousek 1989). Meanwhile, infants with aphasis who have been tracheostomized skip the cooing phase and begin babbling...
(Vihman 2013). Although cooing is one of volitional vocalizations, it may differ from the vocalizations that change to language. Locke pointed out that cooing emerged at the phase when infanticides decreased and played a role in building the vocal interaction between infants and parents (Locke 2006). This vocal interaction with mothers, or “quasi-conversation” experience, might be useful for infants in learning the usage of vocalizations and will also narrow the target of sensory learning.

5.2 Post-cooing Stage

Four-month-old infants begin to vocalize various sounds. They suddenly produce a loud voice and sometimes screech or whisper. Infants often produce these vocalizations without their mothers (Delack 1976). Because they appear to enjoy vocalizing, this is called “vocal playing” (Stark 1980). Through these behaviors, infants may be exploring the articulatory parameters for vocal control and then they begin to learn the articulation. After this stage, the vocalizations shift to babbling. Babbling is a rhythmical vocal sequence that has adult-like sounds with consonants and vowels. This means infants acquire the ability to produce sequential vocalizations with simultaneous control of respiration and articulation. Later babbling includes transition between multiple syllables, rather than the monosyllabic repetition characteristic of early babbling. Infants gradually accumulate pair-wise transition during the later babbling stage (Lipkind et al. 2013). Because of this sophisticated vocal control, it seems that babbling is similar to the sensory-motor learning phase in songbirds.

6. Discussion and Conclusion

We have discussed that learned vocalization, either calls or songs, differ in acoustic and functional characteristics among vocal learners. Songs are vocal trains of acoustically complex syllables with a specific transition pattern. In most cases of song learning, male birds learn songs from their fathers or neighboring territory holders and learning occurs only once in youth. Calls are monosyllabic vocalizations. In call learning, both males and females learn vocalizations. Furthermore, adults maintain a high vocal plasticity and can learn new vocalizations whenever they join a different social group.

Language acquisition in humans has similarities to song and call learners. The acquisition of a native language is limited to the early life stage, just like song learning. On the other hand, humans can learn a second language and there is no gender difference in the ability to acquire language. Equal vocal plasticity between the sexes and vocal plasticity after puberty are similar to what is demonstrated by call learners. Furthermore, human tutor choice, the individual from whom infants learn vocalizations, might be similar to song and call learners. To acquire human language, it is necessary to learn transition patterns or rules that are shared in the social group as well as the vocal sounds. It is not enough to mimic the sounds. Human infants show a preference for their mother’s voice and then toward their native language. These preference shifts suggest there are two types of tutor choices. At the beginning of the sensory learning phase, infants choose their mothers as their tutor and try to mimic the mothers’ vocalizations, just as songbirds do. Subsequently, the vocalization rules that are shared with surrounding people may become the learning target, just as is the case in call learning.

Even non-vocal learners produce sequential vocalizations for eliciting parental behavior and have a primitive sensory-motor learning phase. Before vocal learning, crying to elicit parental behavior may facilitate the development of volitional respiration control in human infants. Vocal development occurs in parallel with sensory-motor learning and is the next learning target. The early vocalizations of human infants have similarities to other vocal learners and non-vocal learners. A quantitative analysis to compare these vocalizations across species would be a valuable next step in advancing the understanding of vocal development.

Acknowledgements

This work was supported by Grant-in-Aid for JSPS Research Fellow 16J40180.

References

Connor, R. C., M. R. Heithaus and L. M. Barre (2001) “Com-
complex social structure, alliance stability and mating access in a bottlenose dolphin "super-alliance.”” *Proceedings of the Royal Society B: Biological Sciences* 268(1464), 263–267.


Stark, R. E. (1980) “Stages of speech development in the first...


Zhang, Y. S. and A. A. Ghazanfar (2016) “Perinatally influenced autonomic system fluctuations drive infant vocal sequences article perinatally influenced autonomic system fluctuations drive infant vocal sequences.” *Current Biology* 26(10), 1249–1260.

（Received Apr. 13, 2017, Accepted Jun. 22, 2017）