

# How vocal temporal parameters develop: a comparative study between humans and songbirds, two distantly related vocal learners

Miki Takahashi<sup>\*,†,‡</sup>, Kazuo Okanoya<sup>§,\*\*</sup> and Reiko Mazuka<sup>\*,†,††</sup>

<sup>†</sup>Laboratory for Language Development, RIKEN Center for Brain Science, Saitama, Japan, <sup>‡</sup>Research fellow of Japan Society for the Promotion of Science, Tokyo, Japan, <sup>§</sup>Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan, <sup>\*\*</sup>Cognition and Behavior Joint Research Laboratory, RIKEN Center for Brain Science, Saitama, Japan and <sup>††</sup>Department of Psychology and Neuroscience, Duke University, Durham NC, USA

\*Corresponding authors: E-mail: reiko.mazuka@riken.jp, miki.takahashi@riken.jp

## Abstract

Human infants acquire motor patterns for speech during the first several years of their lives. Sequential vocalizations such as human speech are complex behaviors, and the ability to learn new vocalizations is limited to only a few animal species. Vocalizations are generated through the coordination of three types of organs: namely, vocal, respiratory, and articulatory organs. Moreover, sophisticated temporal respiratory control might be necessary for sequential vocalization involving human speech. However, it remains unknown how coordination develops in human infants and if this developmental process is shared with other vocal learners. To answer these questions, we analyzed temporal parameters of sequential vocalizations during the first year in human infants and compared these developmental changes to song development in the Bengalese finch, another vocal learner. In human infants, early cry was also analyzed as an innate sequential vocalization. The following three temporal parameters of sequential vocalizations were measured: note duration (ND), inter-onset interval, and inter-note interval (INI). The results showed that both human infants and Bengalese finches had longer INIs than ND in the early phase. Gradually, the INI and ND converged to a similar range throughout development. While ND increased until 6 months of age in infants, the INI decreased up to 60 days posthatching in finches. Regarding infant cry, ND and INI were within similar ranges, but the INI was more stable in length than ND. In sequential vocalizations, temporal parameters developed early with subsequent articulatory stabilization in both vocal learners. However, this developmental change was accomplished in a species-specific manner. These findings could provide important insights into our understanding of the evolution of vocal learning.

## 1. Introduction

Vocal learning is the process of acquiring a new vocal repertoire through experiences (Janik and Slater 2000).

A limited number of animal species learn new vocal repertoires, including five groups in mammals (humans, elephants, pinnipeds, bats, and whales) and three groups in

birds (songbirds, parrots, and hummingbirds). Among vocal learners, humans and songbirds have a common feature in that they can learn a sequence of distinct acoustic notes such as human speech and birdsong (Doupe and Kuhl 1999). Here, we define such vocalizations as ‘sequential vocalizations’. Comparative studies on the learning of sequential vocalizations have revealed similarities between humans and songbirds such as a sensitive period to learn the native language or songs (Doupe and Kuhl 1999), social feedback facilitation for motor learning (Goldstein et al. 2003), and stepwise vocal development (Lipkind et al. 2013). This article, as part of a comparative study, focused on the time-domain developmental pattern of sequential vocalizations in human infants and young songbirds.

To give the reader an idea of how we treat the sequential vocalizations, we describe examples of such vocalizations in a human infant (Fig. 1a) 1) and a species of songbirds, the Bengalese finch (2) (Fig. 1b). In both figures, the upper panel is a sound spectrogram (sonogram) indicating the sound by time (horizontal axis), frequency (vertical axis), and amplitude (gray scale). The lower panel is an amplitude (vertical axis) and time (horizontal axis) plot. When an amplitude threshold is set at a certain level (for a detailed procedure, see the Methods section), it is clear that both vocalizations are organized by the sequence of a note and an inter-note interval (INI). A set of a note and an INI comprises an inter-onset interval (IOI).

Vocal duration is affected by biological constraints such as the vital capacity or breathing rate for gas exchange (MacLarnon and Hewitt 1999). To produce sequential vocalization, which lasts for some duration, exhalation must also be sustained without disturbing the balance of gas exchange. This biological constraint requires sophisticated temporal control of respiratory organs during sequential vocalizations (Lenneberg 1967; Wilder and Baken 1974; Suthers et al. 1999) and could be a primary problem for young vocal learners. There are two possibilities for the initial respiratory control of young vocal learners. One is that respiratory control is provided before the beginning of vocal learning, and the other is that young vocal learners acquire respiratory control during the vocal learning period.

In songbirds, vocal–respiratory coordination is acquired at the beginning of the sensory–motor learning period. Early juvenile songs are asynchronous between song notes and expiration, including expiration without song notes. Through motor practice, juveniles acquire synchronization between phonation and expiration; that is, the onset and offset of each note and expiration match, and the pause between notes matches

inspiration. Clear articulation of notes develops after this pattern matching (Veit et al. 2011).

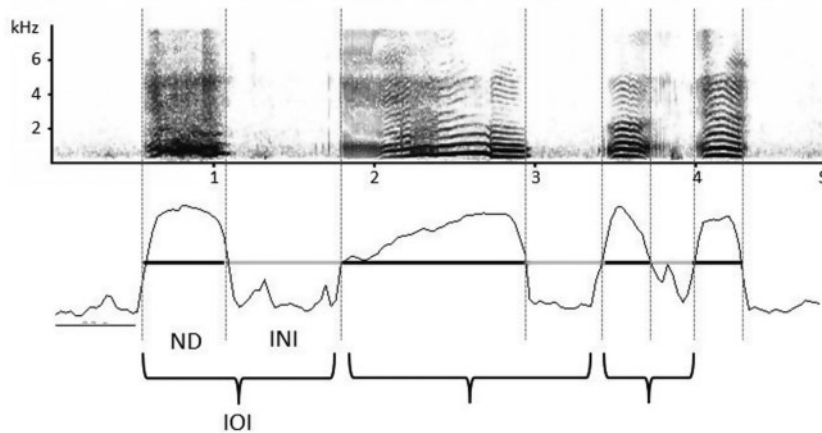
Since the development of human infant vocalization has been studied primarily from the perspective of language development, the development of vocal–respiratory coordination in human infants remains unclear. As reviewed by Vihman (2013), phonological development during the first year has been studied extensively in human infants. Their vocalizations change to harmonic sounds due to an elongated vocal tract with a descended larynx (Lieberman 1984; Fitch and Giedd 1999; Lieberman et al. 2001), and the acoustic features of their vocalizations become more complex as they become attuned to their ambient languages (c.f., Levitt and Wang 1991; Lee et al. 2017).

A few studies analyzing respiratory patterns during vocalizations have shown developmental changes. The expiratory duration of crying increased while the inspiratory duration was stable from 1 to 8 months of age (Wilder and Baken 1974). Developmental changes in respiratory kinematic movements during vocalizations were associated with physiological development (body length or mass) in the first year but not with vocalization types in the second year of life (Boliek et al. 1996; Parham et al. 2011). These results suggest that vocal–respiratory coordination in human infants is strongly related to physiological development and that its development occurs earlier than articulatory learning. However, it remains unclear how developmental changes in the respiratory systems of human infants are reflected in sequential vocalizations rather than in a single syllable or phonation.

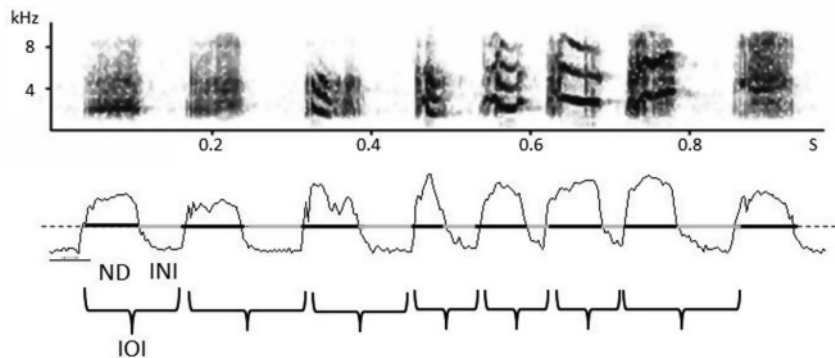
Taken together, these previous studies on songbird and infant respiratory development suggest the early development of the coordination of vocal and respiratory organs, with the subsequent merging of the control of articulatory organs with vocal–respiratory coordination. Considering human infants require almost half a year before vocalization becomes organized into syllable-like units in babbling, the very early stages of human vocalization and songbird vocalization may share similar developmental changes.

However, there is an important difference in their ventilation systems. As shown in Fig. 2a, humans have bilateral lungs. The air goes directly into the lungs and is exhaled from the lungs after gas exchange. Birds feature two major air sacs bilaterally in addition to the lungs, as shown in Fig. 2b. The expansion and contraction of air sacs are responsible for ventilation. Inhaled air enters the posterior air sac first and is then sent to the lungs by contraction of the air sac (Riede and Goller 2010). This system allows birds to breathe more efficiently than

## (a) Human infants



## (b) Bengalese finches



**Figure 1.** Examples of three temporal parameters in sequential vocalizations of a 4-month-old infant and a 63-day-old finch. (a) An example of sequential vocalizations in human infants of 4 month old. (b) An example of songs as sequential vocalizations in Bengalese finches of 63 day old. Top panels are sonograms, middle ones are amplitude envelopes (the ordinate is in an arbitrary unit and the abscissa is time in seconds), which is correspondent to the vocalizations in the top panels. The duration above the amplitude threshold (shown as the straight line) is defined as ND, and the duration below the threshold is defined as INI. IOI is the duration from the onset of the preceding ND to the onset of the following ND that equals to the duration  $ND + INI$ .

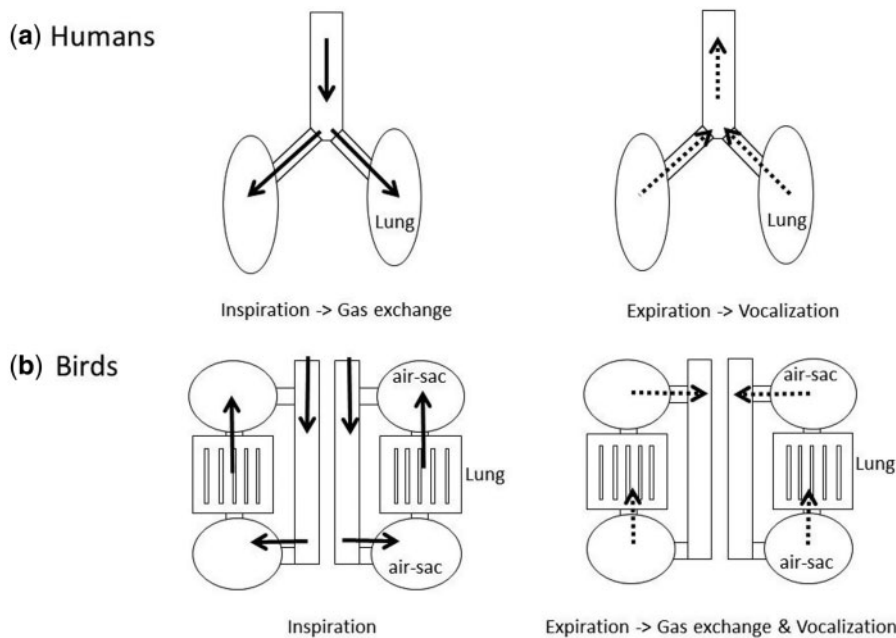
mammals. Thus, human infants and young birds utilize distinct ventilation systems when they resolve the same problem of learning and produce sequential vocalizations.

Here, the question arises whether human infants and young birds have a similar developmental process for vocal learning or different processes reflecting the different ventilation mechanisms. To address this question, we used temporal parameters from sequential vocalizations as respiration-related behavioral measures. It is extremely difficult to directly measure the physiological changes involved in the acquisition of the respiratory control, especially in human infants. The temporal domain of vocalizations is the alternative and

useful measure that can be compared between the two species.

We compared the temporal parameters of sequential vocalizations between human infants and Bengalese finches, a songbird species, as follows: 1) the developmental change of temporal parameters for sequential vocalizations between two species and 2) the temporal parameters in sequential vocalizations and cries in human infants.

As respiratory control is a fundamental factor for vocal production, this would develop earlier than the articulatory progress. Similar developmental changes between infants and young birds could suggest a universal solution for vocal learning. While the early infant cry



**Figure 2.** The respiratory systems of humans and birds. The respiratory systems of humans (a) and birds (b) are different. The human respiratory system requires one cycle (inspiration–expiration) for gas exchange. Oxygen ( $O_2$ )-rich air is introduced into the lungs by inspiration, and gas exchange occurs. Vocalization uses post-exchanged air. In contrast, birds have separated structures for gas exchange and ventilation. Because of this, birds require two cycles for complete gas exchange. The first inspiration introduces  $O_2$ -rich air into the posterior air sac. The following expiration sends air into the lung and, subsequently, gas exchange occurs.

has a form of sequential vocalizations, it is a nonlearning vocalization; that is, it does not require the auditory model to memorize. If the temporal parameters of sequential vocalizations have a similar pattern to the early cry, then it indicates basic vocal–respiratory control might be already provided in early infancy. Comparisons between sequential vocalizations and early crying should identify differences between acquired and nonlearned vocalizations.

## 2. Methods

### 2.1 Developmental periods of vocalizations used for comparison between two species

Infants start to vocalize from the first month. The babblings begin at around 6 months old, and the first word is observed at 1 year old. While it takes several years to learn speech or native language, motor learning to produce articulatory sounds and sequentially vocalize improves during the first year in infants. Therefore, we focused on sequential vocalizations from 0 to 11.5 months of age. In Bengalese finches, young juveniles have motor learning from 35 to 120 days old. Acoustically clear sounds are observed at around 60

days old. The comparison of developmental change between two species was conducted from the beginning to the end of motor learning (0–11.5 months in infants, 35–120 days in finches).

### 2.2 Datasets from human infants

We utilized a published database of Japanese infant vocalizations (Amano et al. 2008). The database includes vocalizations of five Japanese infants (two boys and three girls from three families) longitudinally recorded under natural situations with their parents from the infants' birth until 5 years of age. The recording situation varied, for example, they were alone with the recorder, but at other times they were playing with their mothers. As sequential vocalizations, we selected continuous utterances in which two adjacent vocal elements occurred within 3 s (Gratier et al. 2015) and excluded vocalizations under interactive situations (e.g., turn-taking) to minimize the social effects. Although reflex vocalizations (e.g., laughing and hiccups) were excluded from the analysis, cries from birth up to 1.5 months of age were analyzed for comparison with sequential vocalizations. This is because cries could also be regarded as sequential vocalizations. Crying

acoustically changes through interactive experiences (Yoo et al. 2018). Therefore, we expected that the very early cry as a nonlearned sequential vocalization would make a good comparison with learned vocalizations to find out basic mechanisms for respiratory and vocal control. As described below, cries and other sequential vocalizations were analyzed separately.

### 2.3 Datasets from Bengalese finches

Vocalizations were collected from six male Bengalese finches from six different pairs in a laboratory at Chiba University in 2000. All of the subjects were kept with their parents and siblings until >120 days (full adult) after hatching. Songs were longitudinally recorded once weekly, between 35- and 120-days post hatch. Each recording was made for at least 30 min on one bird in the sound-attenuating box using a digital audio tape recorder (ZA5ES; Sony, Tokyo, Japan).

### 2.4 Analysis of three temporal parameters

We passed the vocalizations of human infants through a high-pass filter at 0.3 kHz to remove ambient noise. However, the recording conditions were not consistent. Background noises (e.g., caregivers' voices, car noises, or toy sounds) and the distance from the microphone differed. Therefore, we normalized the data to 50% of the maximum voltage level. Birdsongs were passed through a band-pass filter between 0.5 and 12 kHz and normalized to 75% of the maximum voltage level before measuring the parameters.

We measured three temporal parameters from sequential vocalizations in human infants and songs in Bengalese finches using automatic parameter measurements provided by SASLab Pro software (Avisoft Bioacoustics, Berlin, Germany) (Fig. 1). Note duration (ND) is defined as the duration of vocalization. The inter-note interval (INI) was defined as the silent period between notes. The inter-onset interval (IOI) was defined as the duration between the onset of the preceding note and the onset of the following note and was, thus, equal to the sum of adjacent ND and INI. These temporal parameters were obtained by setting the amplitude threshold for each recording using SASLab Pro. We determined the threshold level according to each recording condition after excluding the background noise.

The number of vocalizations and recording days differed among infants in the database. We, therefore, searched all recorded data for sequential vocalizations and pooled them by month for each infant. Finally, we analyzed 270 NDs, 238 IOIs, and 238 INIs (range: 36–572) on average per month for sequential vocalizations,

in addition to an average of 209 NDs, 206 IOIs, and 206 INIs of cries (range: 85–276) (Supplementary Table S1). In Bengalese finches, songs from each recording day were analyzed for each bird. An average of 1,371, 1,367, and 1,367 of ND, IOI, and INI per recording day were obtained, respectively (range: 293–3692). Since the recording ages differed between birds, the data were analyzed across subjects by week (e.g., 43–50 days in age) in the following statistical analysis.

### 2.5 Statistical analysis

The first step of the analysis examined the overall distributions of ND, IOI, and INI. The mean and SD of each parameter were obtained by nonlinear regression from the cumulative distributions for each month for the infants' vocalizations and for each day for the finches, using GraphPad Prism version 7.0 (GraphPad Software, La Jolla, CA, USA). Since the distribution patterns were similar to the exponential distribution, using the cumulative distribution was better for comparison of different numbers and orders of duration and ages between infants and birds. The means and SDs were pooled across subjects; subsequently, the developmental change of each parameter was fitted with segmental linear regression to determine a breakpoint. We predicted the temporal domain parameters of vocalization in both human infants and Bengalese finches developed until certain age rather rapidly, and then flatten after that (e.g., Sasahara et al. 2015). Thus, segmental linear regression could show the breakpoint of developmental change.

In addition to the developmental change of the grand average across the subjects, duty cycles were calculated to directly compare data with different orders of duration and age between infants and birds. The duty cycle of ND is the ratio of ND to the sum of the ND and the preceding or following INI (that equals the IOI), while the duty cycle of INI is the ratio of INI toward the same sum.

In the second step, we focused on the adjacent association between ND and INI. The normalized difference between ND and INI was calculated by dividing the difference between ND and INI by the sum of the two (Equation (1)). This normalized difference ranged from –1 to 1.

$$\text{Normalized difference} = \frac{\text{ND} - \text{INI}}{\text{ND} + \text{INI}} \quad (1)$$

If ND and INI had the same duration, then the normalized difference would be 0. If ND was longer than INI, then the difference would be a positive number, while the difference would be negative if INI was longer

than ND. This analysis allowed to know whether the local relationship between ND and INI is similar to the developmental change in the grand averages. For each subject, normalized differences were calculated from the vocal data for each age group. The mean of the normalized differences in each subject was used in the statistical analysis (one-way ANOVA,  $P$  set at 0.05). The preceding and following INIs of each ND were used to calculate the normalized differences.

### 3. Results

#### 3.1 Distribution patterns of parameters

The distribution patterns of the three temporal parameters changed throughout development in both infants and birds. We obtained the best-fitted means and SDs by nonlinear fitting of the distributions of the three parameters by age. Figure 3 shows the developmental changes in the distributions in human infants and birds. In both species, the mean ND was shorter than the mean INI in the initial phase and became comparable in length with age. While infants showed an increase in mean ND, the value was stable in Bengalese finches. The mean infant INI changed less, while that in birds decreased. Developmental changes in the mean IOI seemed to follow the ND change in infants or the INI change in birds. SDs changed similarly to the mean values in both species. The duty cycles also showed the convergence between ND and INI (Fig. 4).

Segmental linear regression was used to determine each breakpoint of developmental change for the means and SDs for ND, INI, and IOI. Figure 3a shows the development of the mean for the three parameters in infants. The breakpoint of the mean ND was 6 months ( $R^2 = 0.97$ ), while that of IOI was 5.5 months ( $R^2 = 0.87$ ). The mean INI changed less than ND and IOI; hence, a horizontal line was more appropriate than a segmental line ( $y = 0.629$ ,  $F_{(3,8)} = 2.6$ ). Figure 3b shows the developmental change in SDs. The breakpoint of SD in ND was 5.6 months ( $R^2 = 0.94$ ), while that of IOI was 6.3 months ( $R^2 = 0.87$ ) (Fig. 3b). The SD of INI had a breakpoint at 5.8 months ( $R^2 = 0.61$ ).

Figure 3c and d show the developmental change of the mean duration and SD of each parameter for birds. The ND did not yield a significant breakpoint in means or SDs in birds and was best-fitted with a horizontal rather than a segmental line;  $y = 0.0637$  for mean ( $F_{(2, 63)} = 1.24$ ),  $y = 0.0227$  for SD ( $F_{(1, 64)} = 1.70$ ). INI showed a breakpoint at 62.6 days for the means ( $R^2 = 0.97$ ) and 62.5 days for the SDs ( $R^2 = 0.73$ ). The IOI means

showed a breakpoint at 59.1 days ( $R^2 = 0.96$ ), and the SDs of IOI ( $R^2 = 0.86$ ) had a breakpoint at 62.5 days.

#### 3.2 Combination of adjacent ND and INI

We calculated the normalized differences between ND and INI to examine whether these factors had similar lengths and to determine their overall distributions. Figure 5a shows the developmental changes in the normalized differences between ND and the preceding INI, while Fig. 5b shows the developmental changes between ND and the following INI in human infants. For human infants, the normalized differences with preceding INIs at 0, 1, and 2 months differed significantly from those of 11.5-month-old infants ( $P < 0.05$ ; Fig. 5a). The normalized differences with the following INIs at 0, 1, 2, and 3 months differed significantly from those in 11.5-month-old infants (one-way ANOVA,  $P < 0.005$ , Fig. 5b; Supplementary Table S2). In finches, the normalized differences both preceding and following INIs at an early phase between 48 and 58 days differed significantly from those at  $\sim 120$  days, the final phase of song learning (one-way ANOVA,  $P < 0.005$ , Fig. 5c and d; Supplementary Table S2).

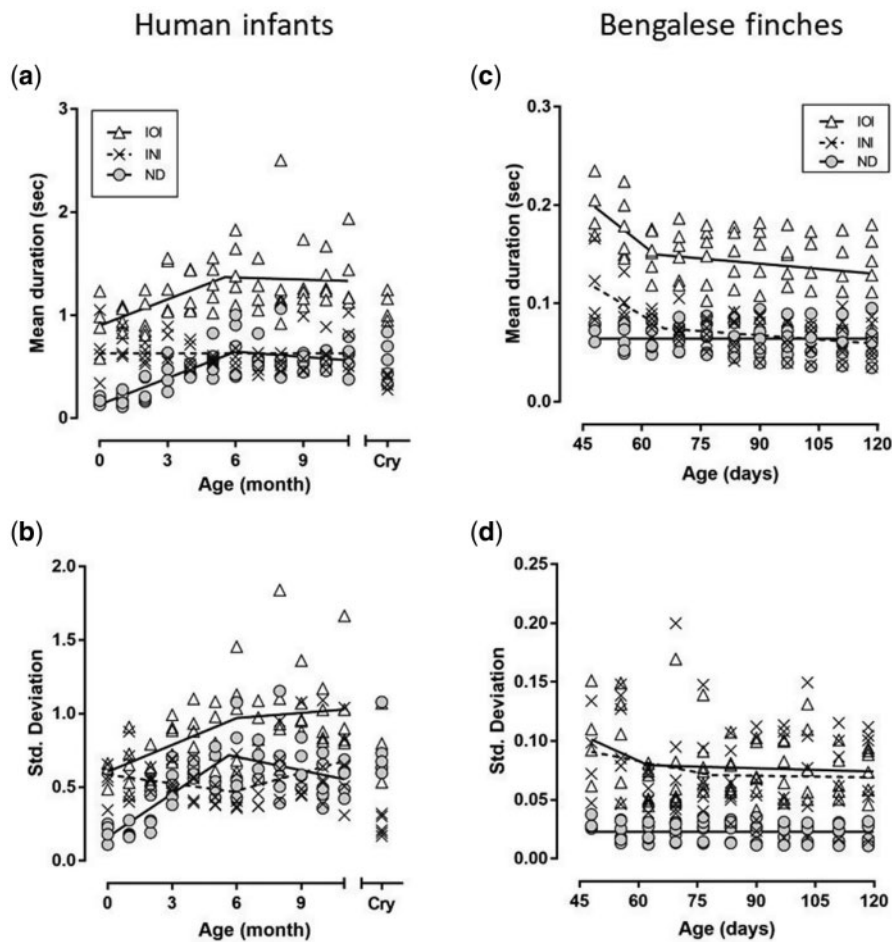
#### 3.3 Comparison between crying and sequential vocalizations in infants

To examine whether crying and other sequential vocalizations differed in terms of temporal parameters, the same three temporal parameters were calculated. The mean and SD for crying are shown in Fig. 3. The distribution of INI for crying fitted within a smaller range compared to that for sequential vocalizations. The distributions of ND and IOI were similar to the range of noncry vocalizations in an older age range ( $>6$  months). The normalized differences for crying were approximately 0, indicating that the ND and INI durations were approximately similar (Fig. 5a and b). Comparisons of the normalized differences between crying and noncry vocalization at 11.5 months showed no significant normalized differences between the two types of vocalizations in either the preceding or following INIs (effect size  $\eta^2 = 0.76$  and 0.73).

### 4. Discussion

#### 4.1 Summary of the results

We compared three temporal parameters from sequential vocalizations during the early phases of vocal learning in two vocal learners: namely, human infants and Bengalese finches. In both species, the mean ND was initially shorter than the mean INI, but they subsequently



**Figure 3.** Developmental changes in means and SD of temporal parameters by nonlinear fitting. Mean duration (a) and SD (b) of three temporal parameters of infant sequential vocalizations from 0 to 11.5 months old, and crying up until 1.5 months old. Crying represented nonlearned vocalizations. (c) mean duration and (d) SD of temporal parameters of the songs of Bengalese finches from 45 to 120 days after hatching. The open triangle is IOI, the X means INI, and the filled circle means ND for each subject. The solid lines for ND and IOI, and the dotted line for INI show the nonlinear fitting with segmental regression.

converged, but the process to attain this differed between the species. In infants, the mean ND was initially shorter than the mean INI but increased until 6 months of age. In finches, the mean INI was initially longer than the ND but decreased until approximately 60 days post hatch. Once the mean ND and INI converged, their mean durations remained stable.

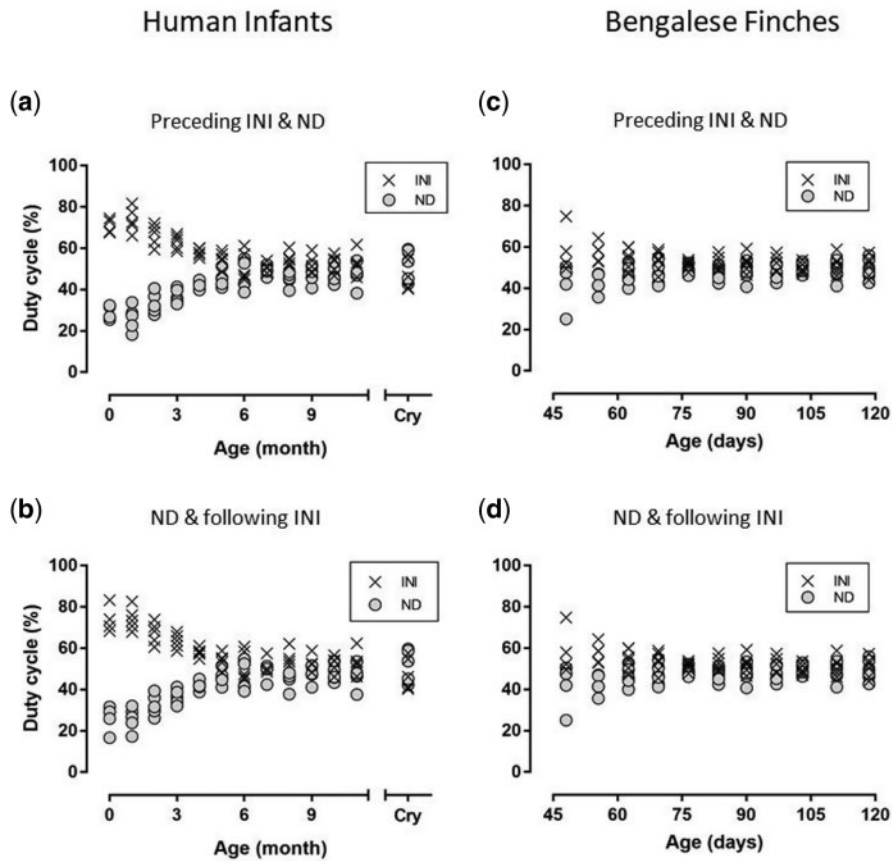
During the early phase of development, the normalized differences between adjacent ND and INI ( $ND - INI$ ) were negative in both species, indicating that the INIs were longer than the NDs at younger ages. However, with age, the difference gradually approached zero.

The results that ND increases while INI stays the same during development are consistent with the prior

work on changes in expiratory/inspiratory activity during crying across ages 1–8 months (Wilder and Baken 1974). However, the dynamics of temporal parameters in human infant cries differed from those in other sequential vocalizations. That is, the mean ND and the INI of cries had similar durations from immediately after birth; thus, the normalized differences between ND and INI were approximately zero from the beginning.

#### 4.2 Differences between human infants and birds

While the convergence between ND and INI occurred in both species, the ND of infants increased and INI of birds decreased. That is, the developmental trajectories of the convergence differed between species. The



**Figure 4.** Developmental change of duty cycles in infants and birds. (a) duty cycles for the preceding INI and ND and (b) duty cycles for ND and the following INI from 0 month to 11.5 month and the early cry in human infants. (c) Duty cycles for the preceding INI and ND, and (d) duty cycles for ND and the following INI from 45 to 120 days in Bengalese finches. The duty cycles of INI are the ratio of INI to the sum of the adjacent ND and INI, and the average for each subject is represented as the symbol X. The duty cycle of ND is the ratio of ND to the sum of adjacent ND and INI, and the mean for each subject is represented as the filled circle.

differences in breathing systems between the species may explain these different developmental trajectories.

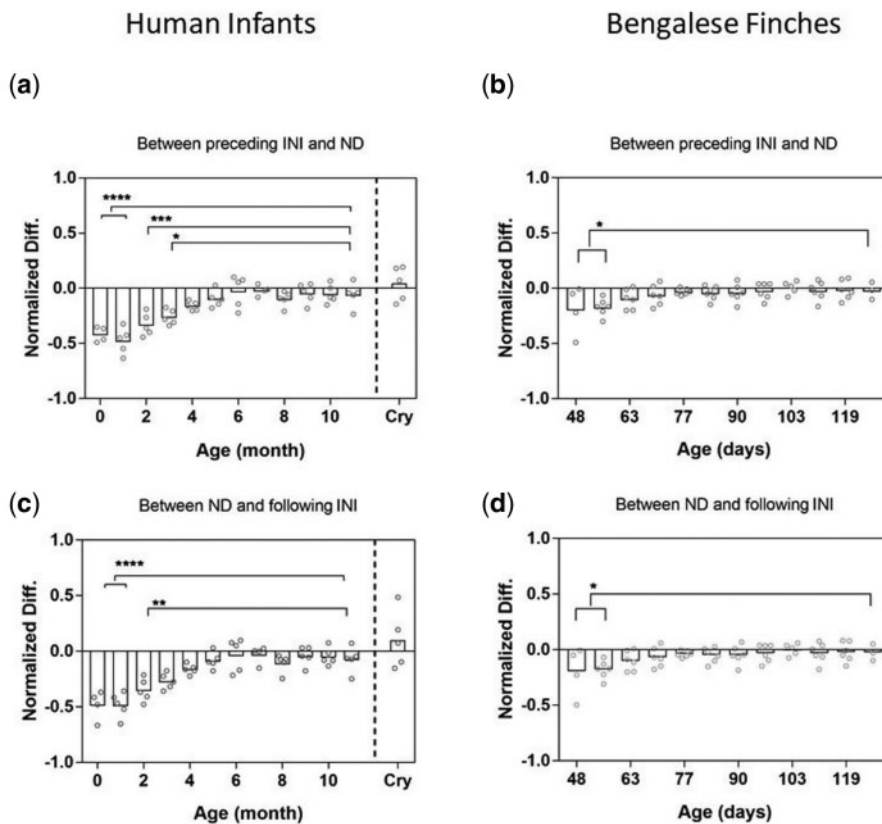
The respiratory systems in each species have distinct characteristics (Fig. 2); while birds have two major air sacs bilaterally besides the lungs, mammals only have bilateral lungs. In birds, oxygen-rich air first enters the air sac, rather than directly into the lungs. As a result, gas exchange and vocalization can occur simultaneously during expiration in birds. However, in mammals, gas exchange follows inspiration and vocalization can occur only during expiration. We hypothesize that in both species, airflow related to gas exchange must be highly accurate to maintain sufficient blood oxygen levels. Therefore, airflow following gas exchange must be stable in duration or timing. On the other hand, airflow after gas exchange could be less constrained and allow more plasticity. This functional consequence of

ventilation dynamics could be manifested in the developmental changes in ND or INI in each species through vocal learning, as the ND of infants increased and the INI of birds decreased.

### 4.3 Similarities between human infants and birds

Humans and songbirds are distantly related and their learned vocalizations are relatively different in terms of acoustic features and hierarchical structure (reviewed by Berwick et al. 2011). However, they share similar patterns of developmental changes in the early phase of vocal learning and temporal parameters as we have shown here. In both species, the initial difference between the mean lengths of ND and INI decreased until a particular age; subsequently, these mean lengths





**Figure 5.** Normalized differences between ND and INI in human infants and Bengalese finches. (a) Developmental changes in normalized difference between preceding INI and ND and (b) between ND and the following INI in human infants. Crying represented nonlearned vocalizations in (a) and (b). (c) Developmental changes in normalized difference between the preceding INI and ND and (d) between ND and the following INI in Bengalese finches. Open circles represent the mean of normalized differences at that age in each subject. Statistical analysis was performed using a one-way analysis of variance ( $P < 0.005$ ) and a *post hoc* test on the data of 11.5-month-old human infants and approximately 120-day post hatch Bengalese finches.

stabilized. This age seemed to be associated with articulation as young Bengalese finches begin to sing with clear acoustic features when they are approximately 60 days old (Sasahara et al. 2015) and human infants begin to babble at approximately 6 months of age (reviewed by Doupe and Kuhl 1999; Vihman 2013). Our results suggest that the temporal parameters of vocalization develop before articulatory motor learning, which may be common among vocal learners, specifically sequential vocal learners.

Why do temporal parameters of sequential vocalization develop earlier than articulation? While multiple reasons are likely, we focus on the following three points. First, vocalization utilizes respiratory airflow. Sequential vocalization requires exhalation without disturbing breathing to maintain adequate oxygen exchange. This restriction is common to all terrestrial animals, including humans and songbirds. The

articulatory pattern must be controlled within the respiratory cycle. Due to this restriction, the temporal pattern of sequential vocalizations associated with the respiratory cycle develops earlier than articulation.

Second, there exist two neural pathways for vocalization in vocal learners. The pathway from the periaqueductal gray (PAG) of the midbrain to the vocal-respiratory center in the medulla is broadly shared among vertebrates (Wild 1997; Jürgens 2002; Kittelberger et al. 2006;) and is responsible for the production of innate vocalizations (reviewed by Jürgens 2009). Newman (2007) argued that the PAG–medulla pathway functions in the sequential vocalization that elicits parental care from adults in altricial animals, including the human infant cry. In addition to this pathway, humans and songbirds have a direct projection from the primary motor area to the vocal-respiratory center in the medulla. This cortico-medullary pathway is

unique to humans and songbirds, which are vocal learners. This pathway is a key feature that enables voluntary control of learned vocalizations (Jürgens 2002). In humans and songbirds, these two pathways connect to the medullar vocal center; one of these pathways, the midbrain–medulla projection, might function as early as immediately after birth (Newman, 2007). As a result, motor control for learned vocalizations merges with the pre-wired PAG–medullary vocal control pathway. Earlier development of the temporal domain might be a common solution for the merging of learned vocal motor patterns with innate vocal motor control.

The third point is related to general motor learning. When a motor sequence such as a reaching movement and serial finger tapping is learned, it is segmented into short motor patterns as chunks. While there are individual differences in chunk formation, the chunk is the foundation of motor sequence learning. A series of chunks reproduces an entire motor sequence (reviewed by Sakai et al. 2004). In line with this, articulation patterns can be considered as short motor units, with several articulatory units embedded into an inhalation–exhalation pair. These units automatically form a chunk. The temporal domain analysis of vocalizations can provide the framework for each chunk within the respiratory cycle.

## 5. Conclusions and future directions

Humans and songbirds have evolved independently to have the same inverse problem in vocal learning: finding an adequate motor pattern that matches sound input. We analyzed the temporal parameters of sequential vocalizations in these two vocal learners and found that short ND and long INI gradually converged until articulation became clear in both species, with ND increasing in humans and INI decreasing in Bengalese finches before convergence. Our results suggest that the learning of sequential vocalizations may have common developmental mechanisms equating the length of INIs and NDs in a species-specific manner, and whether ND or INI adjusted could be caused by species-specific respiratory constraints.

This study focused on developmental changes in only two species of vocal learners. Further comparative studies with nonvocal learners are needed to answer whether the convergence between ND and INI is the foundation for producing sequential vocalizations and whether species-specific development in temporal parameters is caused by different respiratory systems between mammals and birds. Some nonvocal learners also have sequential vocalizations. Among primates, gibbons

(Marshall and Marshall 1976) and indris (Pollock 1986) have territorial songs, and marmosets have babbling-like vocalizations with developmental phonological changes (Elowson et al. 1998; Takahashi et al. 2015). In birds, suboscine species have a long-lasting period for song development even though they are not members of vocal learners (Liu et al. 2013). Comparative studies with these species could be valuable to clarify the remaining questions. We showed that human infants produce long ND during early crying and that the average ND in crying was similar to INI. Infant crying may prepare vocal–respiratory coordination for innate vocalizations. If so, then ND and INI in sequential vocalizations of nonvocal learners might show a converged duration from the beginning.

In conclusion, comparative studies of the developmental changes in the temporal domain of sequential vocalizations between vocal learners might shed light on the origin of vocal learning and how voluntary control for vocalization develops.

## Funding

This work was supported by Grant in Aid for Scientific Research(S) from Japan Society for the Promotion of Science (JSPS) (#16H06319) and Specially Promoted Research from JSPS ((#20H05617) to R.M., by Grant in Aid for Scientific Research on Innovative Areas #4903 Evolving linguistics (#JP17H06380 to K.O. and #JP17H06382 to RM) from the Ministry of Education, Culture, Sports, Science, and Technology, and by Grant in Aid for JSPS Fellows (#16J40180) to M.T.

## Supplementary data

Supplementary data is available at *JOLEV* online.

## Data Availability

Infant vocalizations we analyzed are published as the NTT Infant Speech Database, and it is available from Speech resources Consortium (<http://research.nii.ac.jp/src/en/index.html>). Other datasets for this study including songs of Bengalese finches are available on request to the corresponding author.

*Conflict of interest statement.* None declared.

## References

- Amano, S., Kato, K., and Kondo, T. (2008). Development of NTT infant speech database. IEICE technical report. 108/50,29–34.

- Berwick, R. C. et al. (2011) 'Songs to Syntax: The Linguistics of Birdsong', *Trends in Cognitive Sciences*, 15/3: 113–21.
- Boliek, C. A. et al. (1996) 'Vocalization and Breathing during the First Year of Life', *Journal of Voice*, 10/1: 1–22.
- Doupe, A. J., and Kuhl, P. K. (1999) 'BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms', *Annual Review of Neuroscience*, 22/1: 567–631.
- Elowson, A. M., Snowdon, C. T., and Lazaro-Perea, C. (1998) 'Babbling' and Social Context in Infant Monkeys: Parallels to Human Infants', *Trends in Cognitive Sciences*, 2/1: 31–7.
- Fitch, W. T., and Giedd, J. (1999) 'Morphology and Development of the Human Vocal Tract: A Study Using Magnetic Resonance Imaging', *The Journal of the Acoustical Society of America*, 106/3 (Pt 1): 1511–22.
- Goldstein, M. H., King, A. P., and West, M. J. (2003) 'Social Interaction Shapes Babbling: Testing Parallels between Birdsong and Speech', *Proceedings of the National Academy of Sciences of the United States of America*, 100/13: 8030–5
- Gratier, M. et al. (2015) 'Early Development of Turn-Taking in Vocal Interaction between Mothers and Infants', *Frontiers in Psychology*, 6: 1167
- Janik, V. M., and Slater, P. J. (2000) 'The Different Roles of Social Learning in Vocal Communication', *Animal Behaviour*, 60/1: 1–11.
- Jürgens, U. (2002) 'Neural Pathways Underlying Vocal Control', *Neuroscience and Biobehavioral Reviews*, 26: 235–58
- (2009) 'The Neural Control of Vocalization in Mammals: A Review', *In Journal of Voice*, 23/1: 1–10.
- Kittelberger, J. M., Land, B. R., and Bass, A. H. (2006) 'Midbrain Periaqueductal Gray and Vocal Patterning in a Teleost Fish', *Journal of Neurophysiology*, 96: 71–85.
- Lee, C. C. et al. (2017) 'Subtlety of Ambient-Language Effects in Babbling: A Study of English- and Chinese-Learning Infants at 8, 10, and 12 Months', *Language Learning and Development*, 13/1: 100–126.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.
- Levitt, A. G., and Wang, Q. (1991) 'Evidence for Language-Specific Rhythmic Influences in the Reduplicative Babbling of French-and English-Learning Infants', *Language and Speech*, 34 /3: 235–249
- Lieberman, D. E. et al. (2001) 'Ontogeny of Postnatal Hyoid and Larynx Descent in Humans', *Archives of Oral Biology*, 46/2: 117–28.
- Lieberman, P. (1984). *The Biology and Evolution of Language*. Cambridge: Harvard University Press.
- Lipkind, D. et al. (2013) 'Stepwise Acquisition of Vocal Combinatorial Capacity in Songbirds and Human Infants', *Nature*, 498/7452: 104–8.
- Liu, W. et al. (2013) 'Rudimentary Substrates for Vocal Learning in a Suboscine', *Nature Communications*, 4: 2082.
- MacLarnon, A. M., and Hewitt, G. P. (1999) 'The Evolution of Human Speech: The Role of Enhanced Breathing Control', *American Journal of Physical Anthropology*, 109/3: 341–63.
- Marshall, J. T., and Marshall, E. R. (1976) 'Gibbons and Their Territorial Songs', *Science*, 193/4249: 235–7.
- Newman, J. D. (2007) 'Neural Circuits Underlying Crying and Cry Responding in Mammals', *Behavioural Brain Research*, 182/2: 155–65.
- Parham, D. F., Buder, E. H., and Boliek, C. A. (2011) *Syllable-Related Breathing in Infants in the Second Year of Life*, 54/4: 1039–50.
- Pollock, J. I. (1986) 'The Song of the Indris (Indri Indri; Primates: Lemuroidea): Natural History, Form, and Function', *International Journal of Primatology*, 7/3: 225–64.
- Riede, T., and Goller, F. (2010) 'Peripheral Mechanisms for Vocal Production in Birds - Differences and Similarities to Human Speech and Singing', *Brain and Language*, 115/1: 69–80.
- Sakai, K., Hikosaka, O., and Nakamura, K. (2004) 'Emergence of Rhythm during Motor Learning', *Trends in Cognitive Sciences*, 8/12: 547–53.
- Sasahara, K. et al. (2015) 'A Rhythm Landscape Approach to the Developmental Dynamics of Birdsong', *Journal of the Royal Society Interface*, 12/112: 20150802.
- Suthers, R. A., Goller, F., and Pytte, C. (1999) 'The Neuromuscular Control of Bird Song', *Proceedings of the Royal Society B-Biological Sciences*, 354: 927–39.
- Takahashi, D. Y. et al. (2015) 'The Developmental Dynamics of Marmoset Monkey Vocal Production', *Science (New York, N.Y.)*, 349/6249: 734–8.
- Veit, L., Aronov, D., and Fee, M. S. (2011) 'Learning to Breathe and Sing: Development of Respiratory-Vocal Coordination in Young Songbirds', *Journal of Neurophysiology*, 106/4: 1747–65.
- Vihman, M. M. (2013). *Phonological Development: The First Two Years*. New York: John Wiley & Sons.
- Wild, J. M. (1997) 'Neural Pathways for the Control of Birdsong Production', *Journal of Neurobiology*, 33/5: 653–70.
- Wilder, C. N., and Baken, R. J. (1974) 'Respiratory Patterns in Infant Cry', *Human Communication*, Winter, 18–34.
- Yoo, H., Bowman, D. A., and Oller, D. K. (2018) 'The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-like Vocalizations', *Frontiers in Psychology*, 9/Aug. 1510.