RESEARCH ARTICLE

Filling in the gaps: Acoustic gradation increases in the vocal ontogeny of chimpanzees (Pan troglodytes)

Derry Taylor1 | Guillaume Dezecache2 | Marina Davila-Ross1

1Department of Psychology, University of Portsmouth, Portsmouth, UK
2Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France

Abstract

Human vocal ontogeny is considered to be a process whereby a large repertoire of discrete sounds seemingly emerges from a smaller number of acoustically graded vocalizations. While adult chimpanzee vocal behavior is highly graded, its developmental trajectory is poorly understood. In the present study, we therefore examined the size and structure of the chimpanzee vocal repertoire at different stages of ontogeny. Audio recordings were collected on infant (N = 13) and juvenile (N = 13) semi-wild chimpanzees at Chimfunshi Wildlife Orphanage, Zambia, using focal and ad libitum sampling. All observed call types were acoustically measured. These were predominantly grunts, whimpers, laughs, screams, hoos, and barks and squeaks. A range of spectral and temporal acoustic parameters were extracted, and fuzzy c-means clustering was used to quantify the size and structure of the repertoire. The infant and juvenile vocal repertoires were both best described with the same number of clusters. However, compared to infants, juvenile call clusters were less distinct from one another and could be extracted only when a low level of overlap between call clusters was permitted. Moreover, the acoustic overlap between call clusters was significantly higher for juveniles. Overall, this pattern shows greater acoustic overlap in juvenile vocalizations compared to infants, suggesting a trend toward increased acoustic gradation in chimpanzee vocal ontogeny. This may imply in contrast to humans, chimpanzees become increasingly proficient in using graded signals effectively rather than developing a larger repertoire of more discrete sounds in ontogeny.

KEYWORDS

acoustics, chimpanzee, gradation, ontogeny, vocalization

1 | INTRODUCTION

To understand the evolutionary origins of language, researchers typically search for language-like features in adult primate communication systems (Fitch, 2005). While a great deal has been learned about language origins by adopting this approach, it is crucially limited in that it overlooks the importance of ontogeny. On the ontogenetic path toward language, a key acoustic change in vocal behavior within the first year of life is the emergence of an increasing number of discrete vocal units from a smaller number of more

Guillaume Dezecache and Marina Davila-Ross have equal contributions.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. American Journal of Primatology published by Wiley Periodicals LLC

https://doi.org/10.1002/ajp.23249
acoustically graded vocal units. This transition, widely known as the expansion stage (Oller, 2012; Vihman, 2014), is a routine feature of typical human vocal ontogeny (Stark, 1981) and is accompanied by changes in vocal tract morphology that afford infants a wider acoustic space of possible sounds that they can produce (Crelin, 1987; Lieberman, 1984). Vocal repertoire expansion has been theorized to function as a fitness indicator used to capture the attention of caregivers (Locke, 2017). This is argued to be beneficial in the human ontogenetic niche due to high levels of cooperative breeding and alloparental care (Locke, 2017; Zuberbühler, 2011). Support for this argument comes from studies that show caregivers consider infants with more elaborate vocal behavior to be more likeable (Bloom & Lo, 1990; Bloom et al., 1993) and are more responsive to infants with a more elaborate vocal repertoire (Goldstein & West, 1999).

There is evidence of call types entering and leaving the vocal repertoire during primate vocal ontogeny (Gautier, 1974; Lieblisch et al., 1980). However, whether an expansion in call types occurs at the level of the repertoire is less clear. In cotton-top tamarins, infants and juveniles produce only a single type of long call, whereas adults produce three long-call variants (Snowdon et al., 1983), suggesting vocal-type expansion may occur during the ontogeny of tamarin long calls. Among pigtailed macaques, misclassifications of scream call context based on call acoustics are significantly higher for infant data sets than those from older individuals (Gouzoules & Gouzoules, 1989), suggesting the development of a larger number of context-specific screams. This latter finding is also consistent with vocal-type expansion because juveniles may produce a wider range of scream types than infants. However, it could alternatively be that call types simply become more clearly acoustically differentiated during ontogeny, but call types per se do not expand (i.e., increase in number).

To understand the evolutionary origins of language, it is important to focus on species closely related to humans, such as chimpanzees (Prado-Martinez et al., 2013). Vocal-type expansion is particularly difficult to assess in chimpanzees because the calls of adult chimpanzees are notoriously variable in their acoustic structures, or graded (Slocombe & Zuberbühler, 2010). Despite the gradedness of chimpanzee vocalizations, human listeners have been able to reliably identify particular calls of chimpanzees and link them to distinctive functions (e.g., Goodall, 1990). Moreover, chimpanzee vocal tract morphology shows ontogenetic changes such as descent of the larynx and hyoid bone deeper into the throat that can in principle increase the range of possible sounds that can be produced (Nishimura et al., 2003, 2006). However, it is not known whether repertoire expansion occurs or whether the degree of acoustic gradation of vocalizations changes during chimpanzee vocal ontogeny, as is observed in human vocal ontogeny (Oller, 2012; Stark, 1981).

In fact, chimpanzee vocal ontogeny is a poorly understood process. Some of the earliest descriptions of the vocal repertoire of infant chimpanzees were reported by Frans Plooij in 1984. In a study of wild infant chimpanzees in Gombe National Park, Plooij (1984) classified 22 types of vocalizations in chimpanzees up to 1 year of age based on the researcher’s perception (not based on acoustic analysis). Plooij’s (1984) call classifications mostly consisted of different types of grunts and whimpers, which is consistent with Kojima’s (2008) analysis of infant chimpanzee vocalizations within the first 4 months of life using visual inspection of spectrograms that revealed two broad call categories—grunts and cries. Interestingly, Plooij (1984) did not report any evidence of key features of the adult vocal repertoire such as pant hoot calls, suggesting these emerge later in ontogeny. However, to our knowledge, there are no systematic acoustic studies of vocal behavior focused on stages between infancy and adulthood (i.e., juveniles). Furthermore, all classifications that focus on the vocal repertoire of older individuals (i.e., subadults and adults) vary in estimates of the number of call types between 12 and 32 (Clark, 1991; Goodall, 1990; Van Hooff, 1973). Such varying estimates of the chimpanzee vocal repertoire size create an unclear picture of whether vocal-type expansion occurs during chimpanzee vocal ontogeny and may reflect methodological difficulties in identifying discrete call categories from acoustically graded vocal behavior (Crockford, 2019).

Historically, acoustic gradation in vocal behavior has been largely viewed as a methodological challenge to contend with due to the associated difficulties with identifying reliable vocal units that researchers agree upon (e.g., Crockford & Boesch, 2005; Slocombe & Zuberbühler, 2010). Instead, acoustic gradation can be viewed as a potentially functional feature of a communicative system that increases the information-encoding potential of a signal (Fischer et al., 2017) and may thereby open up new functional possibilities that may be beneficial in more complex social niches. This suggests that by measuring acoustic gradation rather than treating it as noise, researchers can better understand communication (Fischer et al., 2017; Wadewitz et al., 2015). To date though, very few researchers have adopted such an approach.

To our knowledge, such an approach has been adopted in primate research in just one study, which focused on predominantly adult chacma baboons whose vocal repertoire size has been contested in the literature due to high levels of acoustic gradation (see Fischer et al., 2017; Wadewitz et al., 2015). Wadewitz et al. (2015) applied both hard and soft (or fuzzy) clustering techniques to acoustic data extracted from baboon vocalizations. Hard-clustering techniques identify a finite number of discrete clusters (MacQueen, 1967), whereas fuzzy clustering acknowledges that cluster membership is often imperfect and therefore allows a given observation to overlap partially with different clusters (Zadeh, 1965). This cluster overlap can be used to measure gradation. Using hard-clustering techniques, Wadewitz et al. (2015) found that up to 20 call types could be extracted, but the algorithm did not show strong support for any of the sets of clusters that could be extracted, suggesting considerable acoustic gradation in the data. Consistent with this, when adopting a fuzzy-clustering approach, Wadewitz et al. (2015) found that the chacma baboon vocal repertoire was best described as comprising a small number of clusters (two to five) with considerable acoustic gradation between clusters.
In the present study, we aimed to examine ontogenetic changes in repertoire size and acoustic gradation to evaluate whether vocal-type expansion occurs during chimpanzee vocal ontogeny in a sample of 26 young semi-wild chimpanzees, ranging from neonatal infants to juveniles on the brink of sub-adulthood. Leaning on the protocol adopted to study the vocal repertoire of chacma baboons (Fischer et al., 2017; Wedewitz et al., 2015), we assessed the number of clusters that could be reliably extracted, the validity of those clusters, and derived measurements of acoustic gradation (hereafter “typicality”) for both infants and juveniles. If vocal-type expansion occurs during chimpanzee vocal ontogeny as it does in human vocal ontogeny (Stark, 1981), we would expect to observe a higher number of reliable clusters in juveniles, those clusters would be expected to have higher cluster validity, and they would be expected to be significantly less graded.

2 | METHODS

2.1 | Subjects and study site

Subjects were infant (N = 13) and juvenile (N = 13) semi-wild chimpanzees housed at Chimfunshi Wildlife Orphanage, Zambia. Infant ages ranged from 0 up to 4 years of age (M = 1.13 ± SD = 1.14). Individuals aged between 4 and 10 years of age were classified as juveniles (M = 7.07 ± SD = 3.82) (see Supplementary Methods Table S1). These age ranges are the broadest definitions of these developmental stages in the chimpanzee literature and were therefore chosen to give the most comprehensive insight into chimpanzee early vocal development. The infants comprised seven females and six males, whereas juveniles comprised four females and nine males. Typical infant characteristics include riding either ventrally or dorsally with the mother and breastfeeding from the mother. Typical juvenile characteristics include no longer riding with the mother either ventrally or dorsally, less reliance on breastfeeding from mother, and more independence from the mother (see Laporte & Zuberbühler, 2011; Reynolds, 2005; van de Rijt-Plooij & Plooij, 1987). An individuals’ status as an infant or juveniles was, however, defined based on age alone.

Subjects are known to belong to a mixture of subspecies including Pan troglodytes troglodytes and Pan troglodytes schweinfurthii. However, many subspecies memberships were not known, meaning it is possible that the sample contained members of other subspecies. All subjects were raised by their mothers during infancy. All infant subjects still lived with their biological mothers. Three juveniles did not live with their mothers due to fatalities that occurred in years before the present study. We do not believe this impacted the data in the present study because chimpanzee orphans have been shown to differ from others in terms of social behavior and cognition (Beck, 2010) but not vocal acoustics. While the majority of chimpanzees at Chimfunshi are rescued from adverse circumstances such as the pet trade, all subjects in the present study were born in captivity. However, the majority of their mothers were originally wild. Each subject belongs to one of four mixed-sex groups that comprise between 10 and 52 members, including infants, juveniles, subadults, and adults. For a further breakdown of the study population as a function of age, sex, and group, see Supplementary Methods Table S1.

The four groups lived in four outdoor enclosures, respectively. The miombo woodland is their habitat and that of many wild chimpanzee populations (Schoeninger et al., 1999). The enclosures range between 47 and 190 acres in size. Here, chimpanzees are fed once daily at approximately 12 pm, and some chimpanzees are fed indoors. Beyond this, chimpanzees at Chimfunshi may forage on naturally fruiting trees. The enclosures and group sizes are large enough that fission–fusion social dynamics take place as well as regular territory patrols. At night, all chimpanzees sleep outside and many show nesting behaviors.

2.2 | Data collection

Audio recordings were collected between 7 a.m. and 6 p.m. between June 2018 and October 2018 (excluding 12 p.m.–1 p.m. when daily feeding occurred) using a Sennheiser ME66 directional microphone. Recordings were collected only when the subjects were outdoors and the recordist was within 2–10 m of the subject. The main approach in collecting recordings was to use a 5-min focal sampling method in a randomized order each day that allowed us to have an equal representation of the sample in this study. However, due to the large size of the enclosures and the dense forest inside them, subjects were often not visible. For efficiency, we therefore decided to wait for 5 min to observe a subject. If the subject was not visible, we then recorded the visible subject who was next highest on the list. It was attempted to obtain two focal recordings on a subject in a single day—one in the morning (before 12 p.m.) and one in the afternoon (after 1 p.m.). Six hundred and fifty focal recordings were collected overall. Some of these recordings were incomplete as the subjects left the view of the camera for more than 30 s and could therefore no longer be seen and identified as the potential caller. There were 44 incomplete foci. Furthermore, when there were no visible subjects where 5-min focal recordings could be taken for that day, the subjects were also recorded ad libitum. Seventy-nine ad libitum recordings were taken. These additional recordings were also included in the analysis of this study to increase the overall number of calls.

We aimed to obtain 3 h of recordings per subject. Overall, between 15 and 51 focal recordings were collected per subject. Focal observation time ranged between 1.24 and 4.25 h per subject (Infants: M = 2.72 ± SD = 0.96; Juveniles: M = 3.43 ± SD = 0.04). The total duration of ad libitum recordings per subject ranged between 0.03 and 3.52 h (Infants: M = 0.81 ± SD = 0.79; Juveniles: M = 0.18 ± SD = 0.15). The total duration of incomplete focal recordings ranged between 0.20 and 1.27 h (Infants: M = 0.70 ± SD = 0.33; Juveniles: M = 0.55 ± SD = 0.28). Overall, total observation time ranged between 1.73 and 5.45 h (Infants: M = 4.19 ± SD = 1.31; Juveniles: M = 4.16 ± SD = 0.53) per subject.
2.3 | Ethics statement

All data collection was permitted by the University of Portsmouth, Animal Welfare and Ethical Review body (AWERB) and Chimfunshi Research Advisory Board (CRAB). This manuscript also meets the ASP ethical requirements.

2.4 | Identifying calls

The coded unit of vocal behavior was the call type, which is a broad category of calls (i.e., grunts) that contains distinct variants (i.e., food grunt, pant grunt, etc.). Call types were chosen because there is wide agreement regarding the call types produced by young chimpanzees (Goodall, 1990; Plooij, 1984; Slocombe & Zuberbühler, 2010) but whether young chimpanzees exhibit distinct subtypes (i.e., food grunts, pant grunts, etc.) is currently unclear due to a lack of systematic study. Calls would be comprised of a single call element, or a series of call elements otherwise known as a call “bout.” Seven hundred sixty-eight calls were identified in total. The call types included grunts (N = 382), whimpers (N = 147), laughter (N = 139), screams (N = 41), hoo calls (N = 41), barks (N = 8), squeaks (N = 6), and pant hoots (N = 4). Calls were identified based on auditory cues followed by systematic visual inspection of spectrograms according to the definitions in Table 1. The definitions were chosen based on similarities in the definitions of these call types by both pioneering and contemporary primatologists who studied both infants and adults (e.g., Goodall, 1990; Kojima, 2008; Plooij, 1984; Riede et al., 2004; Slocombe & Zuberbühler, 2010). It is important to note that while pant hoots were identified, they were not analyzed because they are calls made up from a series of different call types and are therefore not comparable to the other coded call types. Interestingly, only four pant hoots were observed in the entire observation period, and all were given by juveniles, whereas all other call types were found in both infants and juveniles. For a breakdown of the number of calls produced per subject, see Supplementary Methods Table S2. An inter-rater reliability test was performed on 20% of the total identified calls, and Cohen’s κ revealed a good (see Cohen, 1960) level of reliability (κ = 0.752).

A call started with the onset of acoustic energy in the spectrogram and finished with the offset of acoustic energy in the spectrogram. A bout of elements of the same call type was coded as a single call, unless there was a gap between elements of 4 s or more. If another element occurred after 4 or more seconds or the call type

| TABLE 1 | Definitions of call types based on previous studies of mostly infant and adult chimpanzee vocalizations |
|----------------|------------------------------------------------------------------------------------------|
| Call type | Definition | References |
| Grunt | Short, low-frequency calls given singularly or in short bouts. They may be tonal or noisy and produced with variable rhythm. | Kojima (2008), Plooij (1984), and Slocombe and Zuberbühler (2010) |
| Whimper | Soft low-frequency tonal calls that can become higher in both frequency and amplitude as a bout progresses. Occasionally they may contain harmonics. | Kojima (2008), Plooij (1984), and Slocombe and Zuberbühler (2010) |
| Laughter | Staccato, noisy, low-frequency, alternating ingressive–egressive breathing patterns delivered in an irregular rhythm. Acoustic energy is audibly present in both ingression and egression, with most energy visible during ingresses. While some adopt a more comprehensive laughter definition that includes grunt-like sounds (e.g., Davila-Ross et al., 2009), we decided to adopt a narrower definition with minimal overlap with other call types given the aim of this study. | Plooij (1984) and Slocombe and Zuberbühler (2010) |
| Scream | Loud, high-frequency, harmonic vocalization with varying degrees of tonality (from calls with a high level of noise to calls with clearer harmonics and low-noise levels). Nonlinear phenomena are also commonly observed including frequency jumps (i.e., abrupt changed in fundamental frequency), subharmonics (i.e., spectral components at integer fractional values of the fundamental frequency and as harmonics of these values), biphonation (i.e., two simultaneous but independent fundamental frequencies), and deterministic chaos (i.e., episodes of nonrandom noise). Acoustic energy is usually present only during exhalation, but intense screams also include ingressive sounds. | Goodall (1990), Riede et al. (2004), and Slocombe and Zuberbühler (2010) |
| Hoo call | Tonal call with most energy at onset and a rise and fall in frequency over the call. | Goodall (1990) and Slocombe and Zuberbühler (2010) |
| Bark | Loud calls with abrupt onset. They are often noisy (either due to chaos or turbulence) with a low degree of tonality and are generally low pitched. | Goodall (1990) and Slocombe and Zuberbühler (2010) |
| Squeak | High-frequency, short calls often given in fast succession to form short bouts. These calls are predominantly clear tonal signals. | Plooij (1984) and Slocombe and Zuberbühler (2010) |
| Pant hoot | A call series typically with four distinct phases. Firstly, an introductory phase of low-frequency hoo calls. Second, a build-up phase consisting of increasingly loud panted hoo calls with energy visible on both ingression and egression. Third, a climax phase of screams. Finally, a let-down phase that resembles the build-up phase but was progressively decreasing energy. | Goodall (1990) and Slocombe and Zuberbühler (2010) |
changed, it was considered to be independent of the preceding element. This independence criteria was chosen for comparability with human infant research on this topic that employed a similar criteria (see Oller, 2000).

2.5 Acoustic analysis

Spectral and temporal features were extracted using the bioacoustics analysis programs Raven Pro V1.5 and Praat 6.0.43. Raven Pro was used for extracting all the acoustic data except for formants which cannot be measured using Raven Pro V1.5. In Raven, spectrograms were generated using a fast Fourier transform, whereas spectrograms in Praat were generated using linear predictive coding. Since the majority of extracted measurements were spectral sound characteristics rather than temporal, narrowband spectrograms were chosen. See Table 2 for names of measured parameters and their definitions. A band-pass filter was applied to the spectrograms, ranging from 50 to 20,000 Hz. This bandwidth represented the range of frequencies where energy was visible in previous studies (Mitani et al., 1996; Slocombe & Zuberbühler, 2007). The sampling rate was 48,000 Hz with 16-bit accuracy. A Hanning window function was applied to call selections, which is the most appropriate window function for biological signal analysis because it prevents variation in onset-offset sound characteristics from introducing mathematical artefacts into the acoustic measurements (Clements, 1998). In Praat, identical settings were used as in Raven Pro. See Figure 1 for an example of spectral measurements within a spectrogram.

Call elements were manually selected by highlighting the lowest frequency where there was observable acoustic energy, the highest frequency where there was observable energy, the onset of the call, and the offset of the call. We aimed to highlight the full call in a single selection. When ambient sound from other animals overlapped with call elements of interest, we did not include these features in the selection. Consequently, a single call could comprise several selections. Acoustic measurements were taken for each selection in the call series and the mean average of measurements was taken by dividing the sum total of those measurements by the number of selections taken. The mean value of measurements was not taken for lowest frequency, delta frequency, duration, element number, and element rate. For lowest frequency the lowest observed value across all selections was chosen. Delta frequency was calculated based on the lowest and highest observed frequencies across all selections.

![Figure 1](image-url) Illustration of spectral measurements

| TABLE 2 Acoustic parameters and their definitions |
|-----------------------------------------------|
| Parameter               | Definition                                                                 |
|-------------------------|-----------------------------------------------------------------------------|
| Lowest frequency (Hz)   | The lowest frequency at which energy was detected within the call.           |
| Delta frequency (Hz)    | The highest frequency (the highest frequency at which energy was detected in the call) minus the lowest frequency, corresponding to bandwidth. |
| Peak frequency (Hz)     | The frequency where the highest amplitude value was observed.                |
| Q1 frequency (Hz)       | The frequency below which 25% of the total energy in the selection was found.|
| Q3 frequency (Hz)       | The frequency below which 75% of the total energy in the selection was found.|
| Center frequency (Hz)   | The frequency below which 50% of the total energy in the selection was found.|
| Mean entropy            | Entropy measures the amount of disorder in the selection based on the distribution of energy within the selection. Mean entropy is calculated by first calculating entropy in each frame within the selection and then calculating the mean entropy across frames. |
| 90% bandwidth (Hz)      | The range of frequencies within which 90% of the total energy in the selection was found. |
| F1 (Hz)                 | Formants are resonant frequencies of the vocal tract where acoustic energy is concentrated. F1 refers to the lowest formant in the selection. |
| F2 (Hz)                 | The second lowest formant in the selection.                                 |
| F3 (Hz)                 | The third lowest formant in the selection.                                 |
| Duration (s)            | The amount of time between the onset and offset of a call.                  |
| Element number          | Number of call elements in a call.                                         |
| Element rate (s⁻¹)      | The number of call elements in a call divided by the duration of that call. |
The duration was measured by highlighting the entire call series. Element number was measured by manually counting the number of elements within the selection and element rate was calculated by dividing the element number by the duration of the selection. Since selections were manually taken by highlighting the lowest frequency where there was observable acoustic energy, the highest frequency where there was observable energy, the onset of the call, and the offset of the call, the only parameters that were likely impacted by the selection process were low frequency, delta frequency, and duration. However, stability analyses (see Section 2.6) were performed, which measure whether a small number of parameters have a large impact on the model, which they did not appear to (see Section 3).

2.6 Statistical analysis

Fuzzy c-means clustering was applied to specify the number of clusters of call types in the data, to evaluate the validity of those clusters (how separable they are from one another), the reliability of those clusters (how consistently clusters could be extracted under varying degrees of overlap between clusters), the stability of those clusters (the extent to which cluster differentiation depends on a small number of variables), and to quantify the degree of overlap between clusters (as a measure of acoustic gradation).

Fuzzy c-means clustering cannot process data sets with missing data points. Consequently, we removed call cases where measurements could not be taken, or where some measurements were missing. This resulted in 443 vocalizations, including 221 grunts, 102 whimpers, 83 laughs, 28 screams, 5 barks, 4 hoo calls, and 4 squeaks. Two hundred seventy-three of the calls were infant calls and 188 were juvenile calls. The composition of the infant and juvenile data sets regarding call types were very similar (see Section 3). Since many of the acoustic variables were measured on different scales (i.e., seconds, Hz), before any models were run, all variables were z-transformed to prevent the influence of a parameter on cluster solutions being based mostly on the range of that scale rather than systematic variation within that scale.

z Transformation in particular was chosen because it has been shown to lead to more accurate clustering solutions than other transformation techniques (Mohamad & Usman, 2013), and this technique has also been successfully applied to studies that aim to quantify acoustic call gradation in primates (Fischer et al., 2017; Wadewitz et al., 2015) and birds (Wonke & Wallschläger, 2009).

There are two key parameters that can be controlled in fuzzy c-means cluster analysis K and μ. K specifies the number of clusters to be extracted. μ specifies the degree of overlap between clusters that is tolerated and is known as the “fuzziness parameter.” When μ = 1, extracted clusters are not permitted to overlap. When μ > 1, clusters are permitted to overlap. The greater the value of μ, the more clusters are permitted to overlap. μ can be continuously increased until cluster membership coefficients, which quantify the extent to which a given observation belongs to each cluster and ranges between 1 (complete membership) and 0 (not a member), reach 1/K. At this point, cluster membership coefficients are equal across all clusters and the algorithm can therefore no longer assign cluster membership to observations (Zadeh, 2008; Zhou et al., 2014).

Regarding the K values, we chose a minimum of two clusters to quantify gradation between clusters in the data set, which requires at least two clusters, and a maximum of seven clusters because we included seven call types in the data. Differences between call types in the number of cases included (i.e., only 5 barks compared to 221 grunts) are not problematic in fuzzy c-means clustering wherein there is no minimum requirement for the number of data points per cluster because clustering is based on individual acoustic characteristics of each case rather than the overall number of calls (Gamba et al., 2015; Wadewitz et al., 2015). μ values were systematically varied between 1.1 and 5, at increments of 0.5. This range and increment was chosen because the optimal description of both simulated and real-world data in terms of cluster validity is typically found in the range of μ = 1.1–5 (Zhou et al., 2014). Importantly, previous studies that applied fuzzy c-means clustering to acoustically graded primate vocal behavior found the most stable solution for describing the data had μ = 2 (Fischer et al., 2017; Wadewitz et al., 2015), which is within the range of μ values used in this study.

2.7 Cluster validity and reliability

By systematically varying the K and μ values, 162 solutions were generated, 54 overall solutions, 54 infant solutions, and 54 juvenile solutions. Only infant and juvenile solutions are shown in the main text. All models converged within 500 iterations. The validity of each solution was evaluated based on the mean silhouette value for that solution. Silhouette values represent the degree of confidence that a data point belongs to its primary cluster, as measured by the ratio of mean Euclidean distance between data points within its primary cluster, relative to the mean Euclidean distance between a data point and observations in the nearest neighboring cluster (Bezdek, 1973). Silhouette values can vary between −1 and +1 with values >0 indicating some degree of confidence regarding cluster membership (Bezdek, 1973). Each data point has a corresponding silhouette value, therefore, mean silhouette values were used to measure the confidence of the overall solution. Higher mean silhouette values result from clusters being more distinct from one another.

The range of μ values over which solutions can be computed for any given K value indicates how reliable a model is (Fischer et al., 2017). The value of μ constrains the amount of overlap between clusters, with higher values allowing more overlap. Since the algorithm cannot compute a solution when cluster memberships for that solution are too close to 1/K, if a solution can be extracted only for low values of μ, this suggests the model is not reliable (i.e., robust to overlap between clusters). Models were run and silhouette measurements were extracted in R using the “fanny” implementation of the fuzzy c-means clustering algorithm from the “cluster” package (version 1.15.3) (Maechler et al., 2019).
2.8 | Cluster stability

In addition, fuzzy c-means clustering models can be used to measure the stability of the clusters. Stability measurements compare the results of the fuzzy c-means clustering algorithm when one variable at a time is systematically removed. Using the “clValid” function from the “clValid” package (version 0.6.4) in R, four stability measurements were calculated—mean proportion of non-overlap (APN), mean distance (AD), mean distance between means (ADM), and figure of merit (FOM). A range of measurements was chosen to capture different dimensions of stability, and compare not just overall stability between solutions, but ways in which different solutions are stable. APN measures the mean proportion of data points that change primary cluster membership when the model is systematically recalculated with one variable missing. APN values vary between 0 and 1, with lower values indicating higher stability. AD measures the mean Euclidean distance between observations in the same cluster when the model is systematically recalculated with one variable missing. AD values vary between 0 and ∞ with lower values indicating higher stability. ADM measures the mean Euclidean distance between cluster centers and observations in the same cluster when the model is systematically recalculated with one variable missing. ADM values vary between 0 and ∞ with lower values indicating higher stability. Finally, FOM measures the mean intra-cluster variance of observations in the deleted variable, where clustering is based on the remaining samples. FOM values vary between 0 and ∞ with lower values indicating higher stability. See Brock et al. (2008) for further information on cluster stability measurement in clValid. Together, these analyses measure the extent to which extracted clusters rely on a small number of extreme cases.

2.9 | Cluster composition

After cluster validity and stability was assessed, a hard-clustering solution was extracted for the best-fitting overall, infant and juvenile models. In all cases a two-cluster solution was the best-fitting model and therefore all calls were allocated a value of either 1 or 2 depending on whether their primary cluster membership values were highest for Cluster 1 or Cluster 2. We then cross-tabulated hard cluster membership and call type to examine the contents of each cluster in terms of coded call types. This was done to gain information about the relationship between coded calls and the optimal description of the data based on acoustic information alone, as well as to evaluate the comparability of infant and juvenile call clusters.

2.10 | Cluster typicality

“Typicality” coefficients were extracted, which can be used to quantify the degree of acoustic gradation (Fischer et al., 2017; Wadewitz et al., 2015). Typicality coefficients were calculated by subtracting secondary cluster membership values from primary cluster membership values. Typicality coefficients vary between 0 and 1, with values indicating the percentage of overlap between other clusters. For example, a typicality value of 0.6 indicates 60% of the acoustic space covered by the call overlaps with its primary cluster. Therefore, lower typicality coefficients indicate a higher degree of gradation because this shows more of the acoustic space covered by a call is shared between clusters.

Typicality coefficients were extracted for the best-fitting (defined as the model with the highest silhouette value) overall, infant, and juvenile models, as well as their counterpart models (i.e., if the best-fitting infant and juvenile models had different parameters, infant and juvenile models were created with their own optimal parameters and the optimal parameters for the other data set) in order for values to be comparable across models. For instance, the degree of gradation is highly constrained by the value of the μ parameter because this parameter specifies the amount of overlap between clusters that are tolerated, meaning when μ = 1.1, many more observations would have 100% primary cluster membership compared to when μ > 1.1. Therefore, to make comparisons of typicality coefficients across models, those models must have the same μ values. For each model, typicality was calculated for the whole model to measure the overall amount of gradation, as well as for each cluster, to measure the extent to which clusters were distinct. Typicality coefficients were not normally distributed in any case and were therefore directly compared between data sets using the Mann-Whitney U test to test whether there were significant differences in the degree of acoustic gradation in infant compared to juvenile vocal behavior. For further information on typicality measurements, see Fischer et al. (2017).

The distribution of typical vocalizations was assessed using the approach outlined by Wadewitz et al. (2015). The halved mean absolute deviation of typicality coefficients was calculated. Calls with typicality coefficients greater than this value were classified as “typical” and calls with typicality coefficients lower than this value were classified as “atypical.” This was performed on all models for which typicality coefficients were calculated. This provided a measurement of how many calls are typical of calls in their own cluster, to complement the raw typicality coefficients, which quantify how typical those calls were.

3 | RESULTS

3.1 | Cluster validity, reliability, and stability

For both infants and juveniles, two to seven clusters were successfully identified and two-cluster solutions consistently showed the highest silhouette value (Figure 2). The μ value with the highest silhouette value for infants was 1.5, which generated a mean silhouette value of 0.382 (Figure 2), whereas for juveniles the highest silhouette value was obtained when μ = 1.1, with a mean of 0.304 (Figure 2). This suggests that there is greater overlap between clusters in the juvenile data set compared to the infant data set because silhouette values represent the separability of clusters and μ constraints overlap between clusters, therefore, if the highest
silhouette values occur when the \( \mu \) value is low, this suggests allowing overlap between clusters reduces the separability of clusters, potentially indicating more graded call acoustics.

For infants, two-cluster solutions could be calculated from \( \mu = 1.1 \) to \( \mu = 2.5 \). However, all other solutions could be calculated up to \( \mu = 3.0 \), except the seven-cluster solution, which could be calculated up to \( \mu = 3.5 \) (Figure 2). By contrast, juvenile clusters could only be extracted up to \( \mu = 1.5 \) with the exception of \( K = 7, \mu = 2.0 \). The difference between infants and juveniles in the range of \( \mu \) values over which clusters could be extracted implies that juvenile vocal behavior is more graded than infant vocal behavior, because poor model performance at higher values of \( \mu \) results from a high degree of overlap between clusters. That seven-cluster solutions could be calculated at a larger range of \( \mu \) values suggests that seven clusters could be a more reliable description of the data for both infants and juveniles. However, seven-cluster solutions for infants and juveniles had consistently lower silhouette values than two-cluster solutions and were also less consistent in the silhouette values they generated, which varied from 0.101 to 0.258 for infants (two-cluster solution range = 0.378–0.381) and from 0.160 to 0.221 for juveniles (two-cluster solution range = 0.292–0.304). This suggests that, despite being more reliably extracted, seven-cluster solutions are less consistent in the confidence of call membership per cluster compared to two-cluster solutions.

In comparison to the infant data set, the mean silhouette values for juvenile cluster solutions where \( K > 2 \) did not come as close to the values obtained for models where \( K = 2 \). The mean silhouette value for the next best performing model with \( K > 2 \) for juveniles (\( K = 3, \mu = 1.1 \)) was 0.075 lower than the best performing model. For infants at \( \mu = 1.1 \), the silhouette value for the three-cluster solution was only 0.01 lower than its counterpart two-cluster model. Furthermore, at \( \mu = 3.0 \) a two-cluster solution for infants could not be extracted, but a three-cluster solution whose silhouette value was only 0.003 lower than the best performing model was found. A comparably well-fitting model with a greater than two clusters was not observed for juveniles. This suggests that not only were infant clusters more discrete, but that there was little evidence in the juvenile data set that there could be a higher number of discrete clusters depending on the \( \mu \) values, whereas there was evidence for a third discrete cluster when \( \mu = 1.1 \) and \( \mu = 3 \) in the infant data.

Importantly, while infant and juvenile models differ in both the validity and reliability of clusters, both infant and juvenile models were highly stable. When the models were systematically re-calculated with one parameter missing each time, a mean of only 2.4% of data points changed cluster membership for infants and a mean of 4.5% of data points changed cluster membership for juveniles (see Supplementary Results Section a). This shows that cluster distinctions were not heavily reliant on a small number of acoustic parameters.

### 3.2 Cluster composition

In the best-fitting infant model (\( K = 2, \mu = 1.5 \)), Cluster 1 predominantly consisted of grunts, while Cluster 2 predominantly consisted of whimpers, but also contained a high number of grunts. The majority of all call types except whimpers and hoo calls belonged to Cluster 1 (Table 3). Cluster compositions for the best-fitting juvenile model (\( K = 2, \mu = 1.1 \)) were very similar to that observed in the best-fitting model for infants (Table 3). Cluster 1 predominantly consisted of grunts, while Cluster 2 predominantly consisted of whimpers, but also contained a high number of grunts. Cluster compositions were highly similar for both infants and juveniles with the exception of squeaks, which were found only in Cluster 2 for the juvenile model, while they were found mostly in Cluster 1 in the infant model. However, since only one squeak was observed in the juvenile data set, this is unlikely to indicate any kind of important shift of juvenile vocal behavior. The difference in cluster composition is unlikely to confound infant–juvenile comparisons because these comparisons are made at the level of the cluster rather than the call type. Descriptive statistics on the acoustic characteristics of the best-fitting infant and juvenile models and their counterpart models are shown in Supplementary Results Section b. This generally shows Cluster 1 is comprised of wide-band calls with energy focused at higher frequencies while Cluster 2 is comprised of more low-frequency tonal calls. Cluster composition for the overall model \( K = 7, \mu = 1.1 \) demonstrating the validity of original call coding is shown in Supplementary Results Section b. Descriptive statistics on the acoustic characteristics of all observed call types are shown in Supplementary Results Section g.
Typicality coefficient values are highly constrained by the value of the fuzziness parameter because typicality reflects the degree of overlap between clusters and the degree of cluster overlap is constrained by the value of the fuzziness parameter. Thus, to compare typicality across data sets, typicality was calculated for both the best-fitting model and the corresponding model in another data set, even if that model was not the best-fitting model. It is important to note that there was very little difference in cluster validity between the best-fitting model and the alternative models selected for direct comparison (see Figure 2). As such, the data were comparable within and between ontogenetic stages (i.e., infants and juveniles).

In the best-fitting infant model \((K = 2, \mu = 1.5)\), mean typicality was relatively high \(0.65\), ranging between 0.48 and 0.82 (Figure 3). This pattern was consistent across call types, where mean typicality ranged between 0.62 and 0.81 (see Supplementary Results Section c). In the corresponding juvenile model, mean typicality was relatively low \(0.47\) and coefficients ranged from 0.41 to 0.70 (Figure 3), suggesting higher acoustic gradation in juvenile vocalizations. Mean typicality for juveniles was mostly low across call types, ranging between 0.37 and 0.71 (see Supplementary Results Section c), suggesting higher acoustic gradation among juveniles is not attributable to changes in a single-call type. The suggestion of increased acoustic gradation in juvenile vocalizations was supported by a Mann–Whitney \(U\) test comparing the mean typicality per subject between infants and juveniles, which showed that juvenile typicality measurements were significantly lower than infant measurements \(U = 182, p = 0.0002\). For a further breakdown of infant and juvenile cluster typicality per call type, see Supplementary Results Section d.

In the best-fitting juvenile model \((K = 2, \mu = 1.1)\), mean typicality was high \(0.91\), and coefficients ranged between 0.87 and 0.98 (Figure 3). This pattern was consistent across call types, with

### Table 3

| Cluster number | Call types | Grunt | Whimper | Laugh | Scream | Hoo | Bark | Squeak |
|----------------|------------|-------|---------|-------|--------|-----|------|--------|
| 1              |            |       |         |       |        |     |      |        |
| Infants        | 103 (70.55%) | 12 (17.39%) | 30 (93.75%) | 11 (100%) | 1 (14.29%) | 3 (75.00%) | 2 (66.66%) |
| Juveniles      | 52 (69.33%) | 7 (21.21%) | 40 (78.43) | 17 (94.44%) | 0 (0%) | 1 (100%) | 0 (0%) |
| 2              |            |       |         |       |        |     |      |        |
| Infants        | 43 (29.45%) | 57 (82.61%) | 2 (6.25%) | 0 (0%) | 6 (85.71%) | 1 (25.00%) | 1 (33.33%) |
| Juveniles      | 23 (30.66%) | 26 (78.29%) | 11 (21.57%) | 1 (3.56%) | 8 (100%) | 0 (0%) | 1 (100%) |

**Figure 3** Violin plots depicting the medians, 25th percentiles, 75th percentiles, lower adjacent values, upper adjacent values, and probability density of mean typicality values per individual for infants and juveniles when \(K = 2, \mu = 1.5\) (left) and when \(K = 2, \mu = 1.1\) (right)
typicality coefficients across call types ranging between 0.82 and 0.99 (see Supplementary Results Section e). In the corresponding infant model, mean call typicality was also relatively high (0.95), ranging from 0.80 to 0.99 (Figure 3), suggesting similar call typicality between infants and juveniles in these models. Infant call typicality was also consistently high across call type, ranging between 0.89 and 0.99 (see Supplementary Results Section e). The apparent similarity between infant and juvenile call typicality in these models was supported by a direct comparison between infant and juvenile typicality when $K = 2$, $\mu = 1.1$ using a Mann–Whitney $U$ test which showed that the differences in typicality between these models was not significant ($U = 116$, $p = 0.619$). For a further breakdown of infant and juvenile cluster typicality, see Supplementary Results Section f. Spectrograms of the calls with the highest typicality values for each model are shown below in Figure 4.

**FIGURE 4** Spectrograms of the calls that generated the highest typicality values for each cluster in each best-fitting infant and juvenile model and their counterpart models.

more reliable as the algorithm was able to identify infant clusters over a larger range of fuzziness values compared to juvenile call clusters, indicating less overlap between call clusters in infants compared to juveniles. This inference was supported by analysis of cluster "typicality," which showed that infant call typicality was significantly higher than juvenile call typicality, indicating infant vocalizations are significantly less acoustically graded than juvenile vocalizations.

The observed differences in infant and juvenile call acoustics could be partially attributable to differences in morphology of the vocal apparatus. Magnetic resonance imaging studies have shown that in chimpanzee ontogeny, much like humans, both the larynx and the hyoid bone descend deeper into the throat throughout ontogeny, although not to the same extent as is observed in humans (Nishimura et al., 2003, 2006). This opens up a wider acoustic space of possible sounds that can be produced (Crelin, 1987; Lieberman, 1984), which may therefore represent a physiological mechanism by which the acoustic space between call clusters becomes progressively utilized in chimpanzee ontogeny. However, more recent studies of primate vocal production have shown that even with a larynx positioned high in the throat, an acoustic space of sounds comparable to that of humans is available, leading researchers to argue changes in the diversity of sound production are better explained by neurological changes that underlie vocal control rather than vocal tract morphology (Fitch et al., 2016). This does not, however, address the question of why call types are more acoustically distinct earlier in ontogeny. Signals that are more acoustically consistent are theorized to be more easily interpreted by receivers (McCowan et al., 1999). Since the developmental timing of phenotype emergence can be

4 | DISCUSSION

In the present study, we aimed to examine ontogenetic changes in repertoire size and acoustic gradation to evaluate whether vocal-type expansion occurs during chimpanzee vocal ontogeny. Fuzzy c-means clustering was applied to an acoustic data set derived from the vocalizations of chimpanzees aged 0–10 years. Infant and juvenile repertoires were best described using the same number of clusters, suggesting the vocal repertoire does not become more diverse with respect to the number of discrete call clusters throughout the first 10 years of chimpanzee vocal ontogeny. However, infant call clusters were more acoustically distinct from one another, and were
selected to allow individuals to navigate different ontogenetic niches (Werner & Gilliam, 1984), perhaps less acoustic gradation in the chimpanzee infant vocal repertoire may be a trait that has been selected to help infants navigate their early ontogenetic niche wherein individuals are heavily dependent on caregivers for survival (Plooij, 1984).

While we observed increased acoustic gradation in juvenile vocal behavior, which may be related to their ontogenetic niche, it is also important to consider that the juvenile social world differs from the adult chimpanzee social world. For example, adult males will typically join a core group of dominants and remain within their natal community, while females will often migrate to new communities and spend much more time in small family units (Goodall, 1990). Consequently, to more fully understand chimpanzee vocal ontogeny, it will be essential to quantify acoustic gradation later into adulthood. This will help to better understand the social significance of increased acoustic gradation in juvenile vocal behavior relative to infants, and the role that this plays in subsequent vocal ontogeny. More discrete vocal categories suited to the adult ontogenetic niche may emerge from highly acoustically graded juvenile vocal behavior. Since the adult niche involves different social challenges for males and females, different patterns of vocal ontogeny may also be observed between sexes.

The bioacoustic complexity of signals as measured by diversity in acoustic form also represents a challenge for receivers to contend with because signals with an inconsistent acoustic form are expected to be more difficult to use for communication among living things (McCowan et al., 1999), suggesting increased acoustic gradation may be more challenging for receivers to interpret. Decoding acoustically graded signals may therefore require complex cognitive abilities, and socio-cognitive development in chimpanzees may be related to the increased use of acoustically graded signals. For example, in baboon ontogeny individuals improve in their ability to discriminate between acoustically graded bark variants and respond appropriately to them (Fischer et al., 2000). This example demonstrates that it is important to consider the extent to which vocal behavior is perceived as graded also. Graded primate vocalizations have indeed been demonstrated to be perceived categorically (Fischer, 1998). While two-cluster solutions were most valid in the present study, solutions comprising two to seven clusters for both infants and juveniles that could also be extracted, with some of those solutions outperforming two-cluster solutions with regard to reliability. Neither the most valid nor the most reliable solution is necessarily the most meaningful set of call clusters for the chimpanzee because this depends not only on the acoustic properties of sounds, but how receivers attend to and react to different aspects of those signals. The present study simply identifies a set of possible clusters based on call acoustics and describes the gradation between them. This provides a starting point for subsequent studies to examine which description appears to be most meaningful to receivers based on their responses to calls from different clusters in different models.

The observed pattern stands in contrast to acoustic changes observed early in human ontogeny. Before the onset of canonical babbling, human infant vocal behavior is highly acoustically variable and inconsistent, while later on in the first year of life, between 9 and 12 months of age, the repertoire expands into a larger number of more acoustically distinct vocal categories (Esling, 2012; Goldstein & Schwade, 2008; Mitchell & Kent, 1990). This pattern is predicted by Oller’s (2012) infrastructural natural logic model, which posits that more discrete vocal categories emerge from vocal behavior that is originally more acoustically graded. While there is evidence that in some cercopithecine (Gouzoules & Gouzoules, 1989; Seyfarth & Cheney, 1986) and catarrhine monkey species (Snowdon, 1988) that vocal categories are less acoustically distinct in infancy and become more acoustically distinct later in ontogeny, our observations of chimpanzee vocal acoustics do not follow this pattern. In fact, we have observed the inverse pattern—vocal classes that are originally more distinct early in ontogeny, giving rise to vocal behavior that is more acoustically variable later in ontogeny. Although comparisons between human and chimpanzee communication must be made with caution (see Rendall & Owren, 2013), we might also question the extent to which human speech is comprised of discrete units. Formal linguistics views language as a system with a finite number of discrete elements that can be combined infinitely (Nishimura et al., 2003, 2006). However, when viewed phonetically, natural speech exhibits much acoustic gradation. For example, speech sounds are imbued with prosodic features—acoustic parameters that continuously vary and help to disambiguate lexical meaning by providing information about emotions and arousal (Hammerschmidt & Jürgens, 2007). Additionally, in some languages and dialects, both vowel (e.g., Hickey, 1984) and consonant (e.g., Langacker, 1976; Ulving, 1953) sounds have intermediate forms that grade into one another. Consequently, the degree of gradation in human vocal communication may have been previously underestimated, which could be explored by applying the present protocol to data sets on human vocal acoustics.

Given the aforementioned implications of the present findings, it is necessary to turn our attention to possible methodological issues in the present study. One might ask whether our finding of a lower degree of call gradation in the infant chimpanzees may be the result of the infants showing a propensity to produce call types that have a low degree of gradation more often, instead of having an overall repertoire that is less acoustically graded than the juvenile call repertoire. However, our examination of cluster composition with regard to call type showed that cluster compositions were very similar for infants and juveniles. Cluster 1 for both infants and juveniles was composed of mostly screams, laughs, and grunts, while cluster 2 for both infants and juveniles was composed mostly of whimpers and hoo calls. The proportions of these calls that represented each cluster were also highly similar between infants and juveniles. Therefore, differences in the acoustic gradation of infant and juvenile vocalizations are not explained by a propensity to produce specific call types. Moreover, the observed differences between infants and juveniles in the acoustic gradation also cannot be explained by changes in a single-call type (e.g., just grunts) rather than changes at the level of the repertoire as a whole because typical
measurements were consistently lower for juveniles compared to infants across all call types. Consequently, it appears that the difference in acoustic gradation between infants and juveniles reflects an ontogenetic shift in the acoustic characteristics of the entire repertoire.

One may also ask whether juvenile call typicality was lower because there were fewer calls in the juvenile data set. In fact, we argue that a data set with fewer examples would be expected to appear more discrete because it is less likely to capture the full extent of variation in calls. This is evident when one considers the process of cluster extraction: First, a centroid (the center point of a cluster) is chosen in an $n$-dimensional space (where $n$ = the number of acoustic variables). Next, a new data point is added, and the centroid is recalculated by averaging the data points in that cluster. Clusters are separated by drawing lines in $n$-dimensional space that separate data into the most homogeneous clusters. As the centroids are iteratively recalculated, so too are the $n$-dimensional lines that separate clusters. When there are fewer data points, it is easier to find a line that separates those data points into different groups (Bezdek, 1973; Dunn, 1973). Consequently, when there is a small number of data points, we argue typicality should also be higher because typicality measures how typical a call is of its respective cluster based on the extent to which it overlaps with other clusters compared to the cluster it primarily belongs to, and when there are a small number of data points, overlap between clusters is less likely. Yet, the opposite pattern was found in the smaller data set in this study, which suggests that gradation in the juvenile data set is more likely to be underestimated rather than overestimated because if the juvenile data set was larger, more overlap between clusters would be expected.

To conclude, in the present study, we aimed to quantify ontogenetic changes in repertoire size and acoustic gradation in chimpanzee vocalizations. The results showed that between the infant and juvenile period, the number of call types observed did not differ, but juvenile call acoustics were significantly more graded than infants. This suggests that in contrast to human vocal ontogeny, chimpanzee vocal ontogeny is a process of filling in the acoustic gaps between early call types. Calls with greater acoustic diversity also have greater information encoding potential, raising the intriguing question of how chimpanzees make use of these new communicative possibilities during ontogeny.

ACKNOWLEDGMENTS
We thank D. Kimbrough Oller, Kim Bard, and Sophie Milward for helpful discussions on this topic. We also thank Edwin van Leeuwen and the staff at Chimfunshi Wildlife Orphanage for their support before and during the data collection period as well as Peter Kidd, Abbie Hickinson, and Sarcha Wheeler for help collecting the recordings. Chimfunshi Wildlife Orphanage and Chimfunshi Research Advisory Board gave permission to collect data for this study.

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT
All data and code used for analyses are available upon request.

ORCID
Derry Taylor https://orcid.org/0000-0002-8559-7154
Guillaume Dezecache https://orcid.org/0000-0002-9366-6287
Marina Davila-Ross https://orcid.org/0000-0002-5768-0217

REFERENCES
Beck, B. (2010). Chimpanzee orphans: Sanctuaries, reintroduction, and cognition. In E. Lonsdorf, S. Ross, & T. Matsuzawa (Eds.), The mind of the chimpanzee: Ecological and experimental perspectives (pp. 332–346). University of Chicago Press.

Bezdek, J. C. (1973). Cluster validity with fuzzy sets. Journal of Cybernetics, 3(3), 58–73.

Bloom, K., D’Odorico, L., & Beaumont, S. (1993). Adult preferences for syllabic vocalizations: Generalizations to parity and native language. Infant Behavior and Development, 16(1), 109–120.

Bloom, K., & Lo, E. (1990). Adult perceptions of vocalizing infants. Infant Behavior and Development, 13(2), 209–219.

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R package for cluster validation. Journal of Statistical Software, 25(4), 1–22.

Clark, A. (1991). The socio-ecology of wild chimpanzees vocal behavior in the Kibale Forest, Uganda (Doctoral thesis), University of Michigan, Michigan, USA.

Clements, M. (1998). Digital signal acquisition and representation. In S. L. Hopp, M. J. Owren, & C. S. Evans (Eds.), Animal acoustic communication (pp. 27–57). Springer.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Crelin, E. S. (1987). The human vocal tract: Anatomy, function, development, and evolution. Vantage Press.

Crockford, C. (2019). Why does the chimpanzee vocal repertoire remain poorly understood and what can be done about it? In C. Boesch, & R. Witting (Eds.), The chimpanzees of the Tai forest: 40 years of research (pp. 394–409). Cambridge University Press.

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. Behaviour, 142(4), 397–421.

Davila-Ross, M., J Owren, M., & Zimmermann, E. (2009). Reconstructing the evolution of laughter in great apes and humans. Current Biology, 19(13), 1106–1111.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics, 3(3), 32–57.

Esling, J. H. (2012). The articulatory function of the larynx and the origins of speech. Annual Meeting of the Berkeley Linguistics Society, 38, 121–149.

Fischer, J. (1998). Barbary macaques categorize shrill barks into two call types. Animal Behaviour, 55(4), 799–807.

Fischer, J., Cheney, D. L., & Seyfarth, R. M. (2000). Development of infant baboons’ responses to graded bark variants. Proceedings of the Royal Society of London. Series B: Biological Sciences, 267(1459), 2317–2321.

Fischer, J., Wadewitz, P., & Hammerschmidt, K. (2017). Structural variability and communicative complexity in acoustic communication. Animal Behaviour, 134, 229–237.

Fitch, W. T. (2005). The evolution of language: A comparative review. Biology and Philosophy, 20(2–3), 193–203.

Fitch, W. T. de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. Science Advances, 2(12), e1600723.
Gamba, M., Friard, O., Riodato, I., Righini, R., Colombo, C., Miaretsoa, L., Torti, V., Nadthurou, B., & Giacoma, C. (2015). Comparative analysis of the vocal repertoire of Eulemur: A dynamic time warping approach. *International Journal of Primatology*, 36(5), 894–910.

Gautier, J. P. (1974). Field and laboratory studies of the vocalizations of talapoin monkeys (*Miopithecus talapoin*). *Behaviour*, 51(3–4), 209–271.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants’ babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515–523.

Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, 113(1), 52–58.

Goodall, F. G. (1990). *The chimpanzees of Gombe: Patterns of behaviour*. Harvard University Press.

Gouzoules, H., & Gouzoules, S. (1989). Design features and developmental modification of pigtail macaque, *Macaca nemestrina*, agonistic screams. *Animal Behaviour*, 37(3), 383–401.

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical correlates of agonistic screams. *Animal Behavior*, 73(2), 311–325.

Kojima, S. (2008). Early vocal development in a chimpanzee infant. In T. Matsuzawa (Ed.), *Primate origins of human cognition and behavior* (pp. 190–196). Springer.

Langacker, R. W. (1975). A note on Uto-Aztecan consonant gradation. *International Journal of American Linguistics*, 42(4), 374–379.

Laporte, M. N., & Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. *Developmental Science*, 14(5), 1220–1234.

Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.

Lieblich, A. K., Symmes, D., Newman, J. D., & Shapiro, M. (1980). Development of the Isolation Peep in laboratory-bred squirrel monkeys. *Animal Behaviour*, 28(1), 1–9.

Locke, J. L. (2017). Emancipation of the voice: Vocal complexity as a fitness indicator. *Psychonomic Bulletin & Review*, 24(1), 232–237.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281–297.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2019). cluster: Cluster Analysis Basics and Extensions. R package version 1.5.3.

McCowan, B., Hanser, S. F., & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: Information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, 57(2), 409–419.

Mitani, J. C., Gros-Louis, J., & Macedonia, J. M. (1996). Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. *International Journal of Primatology*, 17(4), 569–583.

Mitchell, P. R., & Kent, R. D. (1990). Phonetic variation in multisyllable babbling. *Journal of Child Language*, 17(2), 247–265.

Mohamed, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299–3303.

Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2003). Descent of the larynx in chimpanzee infants. *Proceedings of the National Academy of Sciences*, 100(12), 6930–6933.

Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2006). Descent of the hyoid in chimpanzees: Evolution of face flattening and speech. *Journal of Human Evolution*, 51(3), 244–254.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Oller, D. K. (2012). Underpinnings for a theory of communicative evolution. In D. K. Oller, & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach* (pp. 49–67). MIT Press.

Plooj, F. X. (1984). *The behavioral development of free-living chimpanzee babies and infants*. Ablex.

Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., Veeramah, K. R., Woerner, A. E., O’Connor, T. D., Santpere, G., Cagan, A., Theunert, C., Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A. E., Malig, M., Hernandez-Rodriguez, J., ... Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, 499(7459), 471–475.

Rendall, D., & Owren, M. J. (2013). Communication without meaning or information: Abandoning language-based and informational constructs in animal communication theory. In U. E. Stegmann (Ed.), *Animal communication theory: Information and influence* (pp. 151–188). Cambridge University Press.

Reynolds, V. (2005). *The chimpanzees of the Budongo Forest: Ecology, behaviour and conservation*. Oxford University Press.

Riede, T., Owren, M. J., & Arcadi, A. C. (2004). Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos. *American Journal of Primatology*, 64(3), 277–291.

Schoeninger, M. J., Moore, J. I. M., & Sept, J. M. (1999). Subsistence strategies of two “savanna” chimpanzee populations: The stable isotopic evidence. *American Journal of Primatology*, 49(4), 297–314.

Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour*, 34(6), 1640–1658.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, 104(43), 17228–17233.

Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. In E. Lonsdorf, S. Ross, & T. Matsuzawa (Eds.), *The mind of the chimpanzee: Ecological and experimental perspectives* (pp. 192–207). University of Chicago Press.

Snowdon, C. T. (1988). Communications as social interaction: Its importance in ontogeny and adult behavior. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 108–122). Springer.

Snowdon, C. T., Cleveland, J., & French, J. A. (1983). Responses to context- and individual-specific cues in cotton-top tamarin long calls. *Animal Behaviour*, 31(1), 92–101.

Stark, R. E. (1981). Infant vocalization: A comprehensive view. *Infant Mental Health Journal*, 2(2), 118–128.

Ulving, T. (1953). Consonant gradation in Eskimo. *International Journal of American Linguistics*, 19(1), 45–52.

van de Rijt, P. (2003). *The stable hyoid in chimpanzees: Evolution of face flattening and speech*. Cambridge University Press.

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A, Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires—Hard vs. soft classification approaches. PLOS One, 10(4), e0125785.

Werner, E. E., & Gillian, J. F. (1984). The ontogenetic niche and species interactions in size-structured populations. *Annual Review of Ecology and Systematics*, 15, 393–425.
Wonke, G., & Wallsläger, D. (2009). Song dialects in the yellowhammer *Emberiza citrinella*: Bioacoustic variation between and within dialects. *Journal of Ornithology*, 150(1), 117–126.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.

Zadeh, L. A. (2008). Is there a need for fuzzy logic? *Information Sciences*, 178(13), 2751–2779.

Zhou, K., Fu, C., & Yang, S. (2014). Fuzziness parameter selection in fuzzy c-means: The perspective of cluster validation. *Science China Information Sciences*, 57(11), 1–8.

Zuberbühler, K. (2011). Cooperative breeding and the evolution of vocal flexibility. In K. R. Gibson, & M. Tallerman (Eds.), The *Oxford handbook of language evolution* (pp. 71–83). Oxford University Press.

SUPPORTING INFORMATION
Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Taylor, D., Dezecache, G., Davila-Ross, M. (2021). Filling in the gaps: Acoustic gradation increases in the vocal ontogeny of chimpanzees (*Pan troglodytes*). *American Journal of Primatology*, e23249. https://doi.org/10.1002/ajp.23249