Title: Paternal aging affects developmental convergence of vocal behavior in individual mice: machine learning-driven analyses of individuality (18/50-word)

Abbreviated title: Novel analyses for USV development in individual mice (46/50-character)

Lingling Mai（麦玲玲）[1], Ryuichi Kimura （木村龍一）[1], Kei Hori （堀啓）[2], Hitoshi Inada （稲田仁）[1,3], Kouta Kanno （菅野康太）[4], Takeru Matsuda （松田孟留）[5], Ryosuke O. Tachibana （橘亮輔）[6], Kaichi Yoshizaki （吉崎嘉一）[1], Valter Tucci[7], Fumiyasu Komaki （駒木文保）[8,9], Mikio Hoshino（星野幹雄）[2] , Noriko Osumi （大隅典子）[1]*

[1]Department of Developmental Neuroscience, Tohoku University Graduate School of Medicine, Sendai, Japan, 980-8575.

[2]Department of Biochemistry and Cellular Biology, National Institute of Neuroscience, NCNP, Tokyo, Japan, 187-8502.

[3]Laboratory of Health and Sports Sciences, Division of Biomedical Engineering for Health and Welfare, Tohoku University Graduate School of Biomedical Engineering, Sendai, Japan, 980-8575.

[4]Faculty of Law, Economics and Humanities, Kagoshima University, Kagoshima, Japan, 890-0065.

[5]Statistical Mathematics Unit, RIKEN Center for Brain Science, Wako, Japan, 351-0198.

[6]Department of Life Science, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan, 153-8902.

[7]Genetics and Epigenetics of Behavior (GEB) laboratory, Istituto Italiano di Tecnologia, Genova, Italy, 16163.

[8]Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan, 113-8656.

[9]Mathematical Informatics Collaboration Unit, RIKEN Center for Brain Science, Wako, Japan, 351-0198.

* Corresponding author, email: osumi@med.tohoku.ac.jp

The number of pages: 56

This PDF file includes:

Main Text

Figure 1 to 6

Tables 1 to 3

Extended data

The number of words:

Abstract: 212, introduction: 570, discussion:927

Conflict of interest statement

The authors declare no competing financial interests.

Acknowledgments

Funding: This work was supported by KAKENHI in Innovative Areas (Grant Number 16H06530) from the Ministry of Education, Culture, Sports, Science, and Technology.

General: Ryuichi Kimura is now at the Department of Drug Discovery Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; Kaichi Yoshizaki is now at the Department of Disease Model, Institute for Developmental Research, Aichi Developmental Disability Center, Aichi, Japan. The authors thank Ms. Sayaka Makino for animal care. The authors also appreciate all members of their laboratories for contributive discussions.

Abstract (212/250 words)

Children with autism spectrum disorder (ASD) show atypical developmental trajectories regarding a variety of phenotypes including vocal communication. Infant crying is a way of social communication and modeled in mice by recording ultrasonic vocalization (USV) induced by maternal separation. Here we applied a novel unsupervised machine learning approach (variational autoencoder, VAE), which can capture comprehensive features of USV to understand the complex phenotypes at an individual level in mice. First, the USVs from neonatal mice derived from *Autism susceptibility candidate 2 (Auts2)* gene mutant fathers were analyzed to validate the VAE. Then, we applied the VAE in a paternal aging model to trace developmental patterns of USV in pups born to aged fathers at the postnatal day 3 (P3), P6, P9, and P12. Our results revealed that pups derived from aged fathers emitted fewer USVs with less diversity and incomplete distribution patterns in an inferring map. While pups derived from young fathers showed a developmental convergence in vocal behavior, those from aged fathers exhibited more atypical patterns with wider variability. Thus, we demonstrate that paternal aging indeed has significant effects on early vocal development and subsequently increases the number of individuals with atypical behavior. The application of VAE also enables us to reveal individual difference, leading to the quantitative research of individuality.

Significance Statement (112/120 words)

Ultrasonic vocalization (USV) in mice is used for a model of vocal communication in autism spectrum disorder (ASD). We mathematically analyzed how maternal

separation-induced USV was impaired in individual pups derived from aged fathers. Our novel approach with unsupervised machine learning (i.e., variational autoencoder, VAE), can detect key features of mouse USVs and identify individual differences during early postnatal stages. This study enables us to describe the individual difference, towards understanding individuality at the quantitative level. Individual variability is crucial in biomedical research to reveal new autistic-like features, investigate the first signs, stratify vocal communication phenotypes in animal models, and develop a translatable biological marker for diagnosing infant crying in the future.

Introduction (570/650 words)

Infant crying is an innate behavior of social communication (Acebo and Thoman, 1992; Acebo and Thoman, 1995) that is used to attract attention from caregivers (Soltis, 2004) and to affect cognitive control in adults (Dudek et al., 2016). Several studies have reported that different crying patterns, such as higher frequency and shorter duration, could be an early sign of a higher risk of neurodevelopmental disorders such as autism spectrum disorder (ASD) (Sheinkopf et al., 2012; Unwin et al., 2017). Therefore, it is essential to establish a systematic procedure for diagnostic analyses of infant crying, leading to further understanding of neurodevelopmental disorders.

ASD is characterized by two core symptoms, i.e., impairment of social interaction (including verbal and nonverbal communication deficits) and repetitive behavior, with various comorbid symptoms (Lotspeich and Ciaranello, 1993; American Psychiatric Association, 2013). Although the etiology of ASD remains unclear, environmental, genetic, and epigenetic factors have been suggested to be intertwined (Schaefer and Mendelsohn, 2008; Tordjman et al., 2014; Nardone et al., 2017). For instance, many genes have been reported to be directly responsible for the symptoms of ASD, such as *FMR1* (Yu et al., 1991), *MECP2 (Amir et al., 1999), SHANK2* (Berkel et al., 2010), and *AUTS2* (Sultana et al., 2002; Bakkaloglu et al., 2008). On the other hand, epidemiological studies have suggested a significant association between father's age and the risk of ASD in offspring (Reichenberg et al., 2006; Hultman et al., 2011).

The ultrasonic vocalizations (USVs) in rodents, especially in mice, have mostly received attention in research related to the neuropathology of communication deficits (Homberg et al., 2016; Jarvis, 2019; Mooney, 2020). Interestingly, separation of a pup from its mother and littermates induces the pup to emit USVs that consist of various sound elements (i.e., syllables), and triggers maternal approach and retrieval behavior (Hahn and Lavooy, 2005; D'Amato et al., 2005 Jan). Thus, maternal separation-induced USVs in mice are considered a form of social communication that corresponds to human infant crying (Esposito et al., 2017). During normal development, USV “syllables” from pups undergo continuous changes in acoustic features and components (Scattoni et al., 2008; Grimsley et al., 2011). Moreover, ASD model mice have been shown to exhibit the differences in various USV parameters, including being fewer in number, and having higher or lower syllable frequency, and shorter syllable duration (Shu et al., 2005; Scattoni et al., 2008; Dougherty et al., 2013; Wohr, 2014). However, USV analyses are commonly based on hand-picked features obtained from manual or semi-automatic methods, and lack a platform to measure individuality. Therefore, the USV analysis remains the challenge in reduced time-consumption, more objectivity, and integrated information.

This study applied a novel unsupervised modeling approach (variational autoencoder, VAE) to comprehensively understand the effects of paternal aging on offspring’s vocal information. Because the *AUTS2* mutation has recently been demonstrated the ASD-like impairments of social interaction and vocal communication in the adult mouse model (Hori et al., 2020), the USVs from the

neonatal pups of *Auts2* mutant mice were first analyzed to validate the method. Then, we investigated the developmental differences of USVs obtained from the pups born to young and aged fathers, paying careful attention to individual variability in syllable patterns. The results show that paternal aging not only quantitatively and qualitatively alters the syllable development at the group level but also impairs the developmental convergence and leads to a gradual increase in the number of offspring with “atypical” USV patterns during early infancy.

Materials and Methods

Experimental animals

1. To create the *Auts2*$^{del8/+}$ pups, *Auts2*$^{del8/+}$ male mice (Hori et al., 2020) of C57BL/6N background were mated with wildtype female mice. Totally eight litters, including 36 wildtype pups and 23 *Auts2*$^{del8/+}$ pups (including both male and female), were obtained and used for USV recording at the National Institute of Neuroscience.

2. To obtain the pups derived from three-month-old (young) or >12-month-old (aged) fathers, C57BL/6J male mice were crossed with 10-week-old (young) virgin C57BL/6J female mice. After mating, each female mouse was separated from the male mouse and isolated to minimize possible confounding factors related to offspring behavior. Totally 32 offspring (21 males, 11 females) from 5 young fathers and 29 offspring (16 males, 13 females) from 5 aged fathers were selected randomly. The average litter size at postnatal day 0 (P0) was not significantly different between YFO (7.00 ± 0.35) and AFO (6.6 ± 0.97) (p = 0.68, two-tailed t-test). The model of paternal aging was created at Tohoku University.

Experimental design

At the scheduled P3, P6, P9, and P12, pups were separated from their mothers and littermates to record the maternal separation-induced USVs (**Figure 1A**). At P3, each offspring derived from young and aged fathers was tattooed using an Aramis Animal Microtattoo System (Natsume Co., Ltd., Tokyo, Japan) for individual recognition after the USV test. Offspring that died during the experimental period were excluded from the analyses (mortality rate, YFO: 5.7% vs. AFO: 12.1%). Finally, the detailed USVs analyses were performed and interpreted. All animals were housed in standard cages in a temperature and humidity-controlled room with a 12-hour light/dark cycle (light on at 8:00) and had free access to standard laboratory chow and tap water. All experimental procedures were approved by the Ethics Committee for Animal Experiments of Tohoku University and the National Institute of Neuroscience. The animals were treated according to the National Institutes of Health guidelines for the care and use of laboratory animals.

USV recordings

According to previously described protocols (Shu et al., 2005; Yoshizaki et al., 2016; Yoshizaki et al., 2017), one by one, each pup was separated from its mother and littermates, placed on a transparent plastic dish with woodchip bedding, and transferred to a sound-attenuating chamber for USV recording. The pups derived from young and aged fathers were recorded 5 minutes after maternal separation at P3, P6, P9, and P12. In comparison, the pups of *Auts2*$^{del8/+}$ and wildtype were recorded 3 minutes after the first detected syllable

during maternal separation at the age of P7. An ultrasound microphone (Avisoft-Bioacoustics CM16/CMPA) was inserted through a hole in the middle of the cover of the chamber, about 10 cm above the offspring in its dish, to record vocalizations. The recorded vocalizations were transferred to an UltraSound Gate 416H detector set (Avisoft Bioacoustics, Germany) at 25–125 kHz. After the recording session, pups were returned to their home cage. This procedure was repeated in sequence until all pups had been recorded. Room temperature was maintained at 22˚C.

Syllable segmentation and classification

Acoustic waveforms were first processed using a GUI-based MATLAB script ("USVSEG") originally developed for segmenting USVs emitted by rodents (Tachibana et al., 2020) because we found that an audio segmentation method implemented in the original variational autoencoder (VAE) detected large amounts of noise. Briefly, the USVSEG computed the spectrograms from each waveform (60 seconds/block), applied a threshold to remove noise from the signal, and detected syllables within a frequency range of 60 –120 kHz in the model of paternal aging and 40 – 120 kHz in the model of *Auts2* mutant. Segmentation criteria for identifying syllables were as follows: a minimum gap of 10 ms should separate two syllables, and each syllable should be a minimum duration of 2 ms. Besides segmentation, the USVSEG automatically calculated the duration, inter-syllable interval, maximum frequency (peak frequency at maximum amplitude), and maximum amplitude of each syllable. Syllable intervals of distinct syllable types were defined as the interval between the type of syllable

of interest and the following syllable. An inter-syllable interval wider than 250 ms was identified as a gap. Segmented syllables were manually classified into twelve categories of syllable types by visual inspection. Ten of these syllable types (#1-10 below) were similar to those described previously (Scattoni et al., 2008; Kikusui et al., 2011). The noise sounds (e.g., scratching noises) that were mistakenly segmented by the USVSEG were manually removed from further analyses. Definitions of each of the syllable categories are as follows (also see **Fig. 1B**):

1) Upward syllables: upwardly modulated with a terminal frequency change ≥ 6.25 kHz compared to the beginning of the syllable.
2) Downward syllables: downwardly modulated with a terminal frequency change ≥ 6.25 kHz compared to the beginning of the syllable.
3) Flat syllables: continuous with a frequency modification ≤ 3 kHz.
4) Short syllables: displayed as a dot and shorter than or equal to 5 ms.
5) Chevron syllables: formed a U or inverted U.
6) Wave syllables: regulated with two-directional changes in frequency > 6.25 kHz.
7) Complex syllables: regulated with three or more directional changes in frequency > 6.25 kHz.
8) One-jump syllables: contained two components with no time interval between them, and the second component displayed a change in frequency ≥ 10 kHz compared to the first component.
9) More-jumps syllables: contained three or more components with no time

interval between them, and the second component displayed a change in frequency ≥ 10 kHz compared to the first and third components.

10) Harmonics syllables: contained one main component stacked on top of other harmonic components of different frequencies.

11) One-jump + harmonics syllables: contained one jump syllable and harmonics syllable with no time interval between them.

12) More-jumps + harmonics syllables: contained a more-jumps syllable and harmonics syllable with no time interval between them.

Characterization of vocalizations

To characterize recorded vocalizations, we used the VAE, an unsupervised machine learning algorithm to extract essential features without any instructional input. We applied the VAE to reduce the dimensions of raw data and quantify subtle changes in behavioral variability on the tens-of-milliseconds timescale to preserve as much information as possible, according to a procedure established previously (Goffinet et al., 2019). The VAE algorithm, which was implemented in PyTorch (v1.1.0), was trained to maximize the evidence lower bound (ELB) standard-setting. The latent dimension was fixed to 32, which the previous study found to be sufficient for all training runs. USV syllables were identified and segmented by USVSEG as described above. After the first training of VAE, a latent map of all syllables was created to remove false-positive syllables as noise syllables. We excluded pups that emitted less than ten calls (2 of YFO and 4 of AFO at P3; 2 of AFO at P6; 2 of AFO at P12; 1 of *Auts2*$^{del8/+}$ pup) from this analysis because data loader creation of VAE requires a minimum of 10 syllables.

After removing noise, the VAE was trained to map single-syllable spectrogram images, such as those shown in **Fig. 1C**. T-distributed Stochastic Neighbor Embedding (t-SNE) was applied for each pup to visualize variations among individuals (**Fig. 2G** and **Fig. 6B-5C**). The distance between points represents the similarity in vocal repertoire.

Statistical Analysis

For better visualization, the acoustic features of syllables are presented as Z scores: (score – mean) / standard deviation. In the model of paternal aging, a two-way analysis of variance (ANOVA) with false discovery rate (FDR) correction (α = 0.05) was used to investigate the statistical significance of differences in syllable parameters, including Z scores of the number, duration, maximum frequency, maximum amplitude and interval of overall and distinct syllables, number of syllable types, and entropy scores. The ANOVA examined two main effects (i.e., father's age and postnatal day) and their interaction (i.e., father's age × postnatal day). In the model of *Auts2* mutant, the Wilcoxon test was performed to detect the differences of Z scores of the number, duration, maximum frequency, maximum amplitude, interval of overall syllables, number of syllable types, and entropy scores between the two groups. A multivariate analysis of variance (MANOVA) was performed to detect group differences in the syllable percentage component, principal components, and t-SNEs at each postnatal day with the father's age or genotype as the independent variable. *Post-hoc* comparisons were performed using a two-tailed *t*-test with Benjamini-Hochberg correction (BH correction) in the syllable percentage component.

To detect the diversity of syllable types, we used the information entropy to measure the uniformity in production rate across syllable types for each offspring. The entropy score ranged from 0 to 1. The score approaches 1 when an animal uniformly (or diversely) produces all of the syllable types and approaches 0 when an animal prefers to produce fewer specific syllable types (less diverse). The entropy score was calculated using the following equation:

$$\text{entropy} = \frac{-\sum_{t=1}^{T} p_t \log_2 p_t}{\log_2 T}$$

Where T indicates the number of syllable types, and p_t indicates the production rate of a specific syllable type t. We excluded offspring (2 of YFO and 4 of AFO at P3; 2 of AFO at P6; 2 of AFO at P9; 2 of AFO at P12: 1 of *Auts2*$^{del8/+}$ pup), whose total number of syllables was insufficient (less than 10) for analysis of entropy. Entropy scores and number of syllable types were compared between YFO and AFO across different postnatal days using two-way ANOVA and tested between wildtype and *Auts2*$^{del8/+}$ pups by Wilcoxon test. Clustering analysis with Gaussian mixture models (GMMs) was applied to understand the developmental trajectories at the individual level, where the data dimension of eight corresponded to the number and duration of syllables at four-time points. We fit GMMs with diagonal covariance Gaussian components using the MATLAB function *fitgmdist*. The number of clusters was determined by minimizing the Akaike information criterion (AIC) (McLachlan and Peel, 2000; Konishi and Kitagawa, 2008). Based on the fitted GMM, each mouse pup was categorized into the cluster with maximum posterior probability. The chi-squared independence test was subsequently applied to determine whether the cluster

distribution was significantly different between the two groups. Principal component analysis (PCA) was performed to objectively characterize the syllable patterns of individual offspring. In the present study, syllable data, including syllable number, number of syllable types, duration, maximum frequency, and maximum amplitude, were normalized by Z score and inputted for PCA to generate principal components. For all comparisons, the significance level was set at $\alpha = 0.05$. JMP13 Pro software (SAS Institute, Cary, NC, USA) was used for statistical analyses. Values are shown as mean ± standard error of the mean (S.E.M.) for each group. Scatter plots shown in **Fig. 2A, 2C-D,** and **2F-G, Fig. 6A–B** were constructed using JMP13 Pro software, and **Fig. 6C** was made using R software.

Results

The USV impairments in *Auts2*$^{del8/+}$ pups

We first analyzed the USVs from pups of *Auts2*$^{del8/+}$ mouse, an established genetic model of ASD (Hori et al., 2020), at postnatal day 7 (P7) to validate procedures of VAE analyses by comparing it to a conventional approach. As the preprocessing step of the VAE analysis, datasets of USVs recorded for three minutes after the first syllable's capture from the wild type and *Auts2*$^{del8/+}$ littermates were automatically characterized by the USVSEG to obtain acoustic parameters of syllables (**Fig.2A, Table 1**). A reduced number, shorter duration, and lower amplitude of syllables, along with longer intervals between adjacent syllables, were observed in *Auts2*$^{del8/+}$ pups. Then, we took a conventional approach by categorizing all syllables into twelve types based on their

spectrotemporal patterns, according to previous reports (Scattoni et al., 2008; Kikusui et al., 2011; Hori et al., 2020). After syllable classification, the significantly different syllable compositions of wildtype and *Auts2*$^{del8/+}$ pups were also detected (F (11, 47) = 4.483, $p < 0.001$, MANOVA; **Fig. 2B and Table 2**). Compared to the syllables from wildtype pups, a higher percentage of "flat" and "short" syllables, together with a lower percentage of "chevron," "wave," "complex," "one-jump," "more-jumps," "one-jump + harmonics," and "more-jumps + harmonics" syllables emitted by *Auts2*$^{del8/+}$ pups. To evaluate the diversity of syllable types, we calculated entropy scores as an indicator of production uniformity across the syllable types in individual offspring. As expected, the number of syllable types and the entropy scores showed significant reductions in the syllables from *Auts2*$^{del8/+}$ pups (**Fig. 2C** and **2D**). Following the USVSEG, VAE performed the training and mapped the syllables of wildtype and *Auts2*$^{del8/+}$ pups in an inferring space (**Fig. 2E**). In this map, the syllable distribution patterns were generally similar between the pups of wildtype and *Auts2*$^{del8/+}$. However, some areas of syllables, which highly concentrated in the syllable map of the wildtype pups, was totally vanished in that of *Auts2*$^{del8/+}$ pups.

To detect the differences at the individual level, PCA from the USVSEG and t-SNE from the VAE were performed and visualized, respectively (**Fig. 2F** and **2G**). At P7, both PCA (F (1, 57) = 4.304, $p = 0.043$, MANOVA) and t-SNE (F (1, 56) = 4.034, $p = 0.049$, MANOVA) demonstrated a significant distinction of vocal behavior between wildtype and *Auts2*$^{del8/+}$ pups, even though some of the pups in the two groups shared an overlapping region. Therefore, the autistic-like

USVs induced by loss of *Auts2* were clearly observed at P7 through the VAE analysis.

Paternal aging alters the acoustic features of syllables emitted by offspring

After the validity of VAE has been tested, the effects of paternal aging on USVs across four early postnatal stages were investigated. Offspring were obtained by mating young wildtype female mice with either young (3 months) or aged (>12 months) wildtype male mice (YFO and AFO, respectively). A large number of syllables were collected from the YFO and AFO at P3, P6, P9, and P12 (**Table 1**). The USVSEG output data showed that paternal aging altered the trajectories of overall syllables emitted by the AFO, as reflected by a reduced number and shorter duration of syllables and different intervals between adjacent syllables (**Fig.3A**).

The syllables were subsequently classified into twelve types to reach a deeper understanding. The USV acoustic features of the twelve syllable types are shown in **Fig. 4**. In general, a total of nine types of syllables (upward, downward, short, chevron, wave, complex, one-jump, more-jumps, and more-jumps + harmonics) were significantly altered in the AFO. Among these altered syllables, six types (upward, chevron, wave, complex, more-jumps, and more-jumps + harmonics) showed a reduction only in number. Changes to two features were detected in the “downward” syllable (shorter duration and different developmental trajectory of intervals) and the “short” syllable (decrease in the number and lower maximum frequency). The “one-jump” syllable exhibited

changes to four features (reduction in number, shorter duration, the lower and different trajectory of maximum frequency, and different trajectory of intervals). Therefore, paternal aging had varying levels of effect on distinct types of syllables.

We conducted further analysis of the composition of the twelve syllable types (**Fig. 3B** and **Table 3**). While no statistical difference was detected at P3, the AFO emitted USVs with significantly different syllable components to those of the YFO from P6 to P12. Compared to the syllables from the YFO on each postnatal day, the AFO USVs comprised a higher percentage of “downward” syllables at P9 and P12. By contrast, other minor syllables such as “chevron,” “wave,” “complex,” “one-jump,” “more-jumps,” “one-jump + harmonics,” and “more-jumps + harmonics” made up a lower percentage of USVs emitted by the AFO during postnatal development. These results show that paternal aging led to significantly different compositions of distinct syllables.

Next, we focused on the diversity of syllables. We found that the YFO emitted USVs containing more types of syllables along with age, while the AFO produced USVs with significantly fewer syllable types from P3 to P12 (**Fig. 3C**). Correspondingly, the YFO emitted syllables with increased entropy scores during the postnatal period, consistent with a previous study (Grimsley et al., 2011), while the AFO showed lower entropy scores than the YFO across all postnatal stages (**Fig. 3D**). These data indicate that the AFO exhibited fewer types and a narrower diversity of syllables, thus had a poorer vocalization repertoire.

To characterize and quantify subtle changes in USVs, we next applied the VAE and mapped a series of points in the inferred latent space (**Fig. 3E**). The number of syllables detected by the VAE again demonstrated that the AFO emitted a smaller number of syllables (father's age, $F(1, 233) = 6.854$, $p = 0.028$, ANOVA) (**Table 1**). Moreover, a different distribution of the syllables at each developmental stage in both the YFO and AFO showed a dynamic process of USV development during the early postnatal stages. Interestingly, differences between the YFO and AFO became apparent from P6 and were particularly evident at P12, although at P3 syllable distribution patterns were similar between the two groups. The significant difference was the presence of a distinct aggregated pattern in USVs emitted from the YFO, which was less obvious in the AFO. Missing parts of the aggregated syllables may be explained by the reduced number and lack of important syllables that cannot be developed in some atypical individuals. Thus, the VAE confirmed the syllable composition and entropy results from USVSEG and demonstrated, without bias, that paternal aging indeed led to deficiencies in the USV dynamic development of composition and diversity.

Paternal aging increases the number of offspring with atypical syllable patterns

We further examined individual differences between the YFO and AFO because children with ASD often exhibit various atypical developmental trajectories. To understand longitudinal syllable development in individuals, we performed clustering analyses based on the USVSEG data using Gaussian mixture models.

The Akaike information criterion (AIC) was applied to determine the cluster number.

Because the number and duration of the overall syllables were positively correlated (YFO: Pearson's correlation coefficient 0.519, $p < 0.001$; AFO: 0.511, $p < 0.001$) and syllables emitted by the AFO were significantly decreased in number and duration, we first separated the offspring into clusters based on syllable number and duration (**Fig. 5A**). The AIC suggested that five clusters were optimal. The developmental patterns in syllable number and duration observed in the YFO were distributed across the five clusters and concentrated in the fourth cluster. In contrast, those in the AFO were only distributed among four clusters and concentrated in the third cluster. Chi-squared independence tests revealed that the clustering pattern (i.e., the proportion of individuals in each cluster) was significantly different between the YFO and AFO ($p = 0.02$). These results show that the dominant developmental patterns of the syllable number and duration were different between the YFO and AFO.

We next classified the number of syllable types into four clusters based on the AIC (**Fig. 5B**). Syllables emitted by the YFO and AFO were classified into three and four clusters, respectively. Most of the YFO belonged to the second cluster, while most of the AFO belonged to the first cluster. Interestingly, the third cluster contained only individual AFO. Cluster distribution was significantly different between the two groups ($p = 0.003$, chi-squared test). Moreover, to examine the development of syllable diversity in individuals, we clustered normalized entropy (**Fig. 5C**). Five clusters were used based on the AIC. Cluster

distribution was significantly different between the YFO and AFO ($p < 0.001$, chi-squared test). The YFO occupied four clusters and dominated the fourth cluster, while the AFO occupied five clusters and dominated the first cluster. Again, the second cluster comprised only individual AFO. These findings indicate that the developmental trajectory of syllable diversity in the AFO differs from that in the YFO. Therefore, the clustering analyses uncovered a unique pattern without obvious gender bias in each pup and the influence of paternal aging in the longitudinal developmental patterns of the number and duration, the number of syllable types, and diversity.

To summarize the individual patterns across transversal stages, we applied PCA analyses to USVSEG data and t-SNE analyses to VAE, respectively, to group the syllable features exhibited by each individual (**Fig. 6**). In the PCA analyses (**Fig. 6A**), at P3, clusters comprising data from 90% of offspring in each group were large, indicating wide individual differences in vocal communication between pups regardless of paternal age. However, there was a minor but significant difference between the YFO and AFO ($F(1, 59) = 5.348$, $p = 0.024$, MANOVA). This difference was more obvious from P6 to P12 (P6, $F(1, 59) = 10.353$, $p = 0.007$; P9, $F(1, 59) = 8.512$, $p = 0.004$; P12, $F(1, 59) = 21.243$, $p < 0.0001$, MANOVA). While the YFO showed developmental convergence in USV patterns, the variability among the AFO remained wide even at P12. In the t-SNE analyses (**Fig. 6B**), again, the vocal repertoires of the YFO demonstrated developmental convergence despite different trajectories, whereas those of AFO were widely scattered during development. The most significant difference was

observed at P12 (F (1, 57) = 25. 462, $p < 0.0001$, MANOVA), which is consistent with the PCA results. No gender bias was observed in this phenotype. We subsequently outlined the data from individual traces and highlighted the initial stage (P3) and terminal stage (P12) (**Fig. 6C**). In both the PCA and t-SNE, the inferred spatial positions of YFO and AFO were similar at P3 but became significantly different in later developmental stages. Thus, paternal aging gradually led to the mis-regulation of vocal development and increased pups with atypical vocalizations.

Discussion (927/1500 words)

The VAE represents the first attempt to derive unsupervised data-driven developmental trajectories from maternal separation-induced vocalizations in pups. Through overcoming the limitations of conventional approaches such as failing to capture variations and to uniform the standards across laboratories, the VAE learns and compresses the data into a small number of latent variables without bias (Goffinet et al., 2019). It demonstrates the capacity of best capture variability and quantifying vocal behavior that provides us novel comprehension into the maternal separation-induced USVs. Currently, with advances in machine learning technology, big data from over 20,000 or 140,000 people were used to analyze individual personality in humans (Gerlach et al., 2018; Nave et al., 2018). However, these machine learning approaches have used "supervised" learning and the results were still similar to the twenty-century paradigm (Goldberg, 1990). In comparison, without setting restrictive parametric assumptions, the unsupervised approach, VAE can characterize the variations of every single-

syllable and individual as latent features that visualize the relationships among elements. These new features obtained from the VAE provide us a novel insight that is difficult to be explored by the traditional paradigm.

Maternal separation-induced USV in mice is considered to be a model corresponding to human infant crying (Branchi et al., 2001; Scattoni et al., 2009). There are accumulated literatures showing USV impairments in genetic models of ASD such as *Cd157* knockout mice (Lopatina et al., 2017), *Tbx1*-deficient heterozygous mice (Takahashi et al., 2016), *Reelin* mutant mice (Romano et al., 2013), and *fmr1* knockout mice (Lai et al., 2014). However, all these findings were based on conventional analyses. To obtain deeper understanding of USV phenotypes, we have applied a novel unsupervised approach using the VAE together with the semi-automatic USVSEG to a genetical model, i.e., *Auts2* mutant mice. We found besides the ASD-like impairments in the traditional features such as the number of syllables, duration, amplitude, syllable diversity, and syllable component, the VAE uncovers the missed areas in a learned syllable latent map and different variations of individuals in *Auts2*$^{del8/+}$ pups at the postnatal day 7. The VAE allows us to conceptualize the USV deficiencies through mapping every single-syllable and recapitulating individuals. These interesting results encouraged us to further examine more developmental stages in the paternal aging model to quantify the subtle variability and comprehend individuality during development.

Advanced paternal age is a risk factor for ASD (Reichenberg et al., 2006; Hultman et al., 2011; Sandin et al., 2016; Kojima et al., 2019), which is modeled

in mice by several laboratories, including us (Sampino et al., 2014; Janecka et al., 2015; Yoshizaki et al., 2021). Since atypical behaviors with diverse heterogeneity are important features of ASD (American Psychiatric Association, 2013; Grzadzinski et al., 2013; Wozniak et al., 2017), depicting individuality could capture the variation across individuals and avoid missing subtle differences (Szatmari et al., 2007; Loth et al., 2017). However, many previous studies have ended with only the differences in USV numbers between groups, and no attention has been paid to systematic analyses of the individual difference during postnatal development. The novel approach we presented here enables us to appreciate the neurobiological significance of vocal behavior in individuals without bias.

Here, the VAE revealed a wider variability and “atypical” USV development among individuals of the AFO. Our results showed that the AFO emitted fewer USVs with less diversity and incomplete distribution patterns on the inferring map. In addition, the YFO showed a typical developmental pattern that gradually converged with age and displayed a relatively uniform behavior at P12. On the other hand, in the AFO, more varied and atypical patterns were observed, indicating that paternal aging has a significant effect on early vocal development, and the number of individuals with atypical behavior increases sequentially.

Interestingly, our data demonstrate that the stage around P6 may be a critical window in which ‘atypical’ patterns begin. This is because the USV features of YFO individuals began to converge, while differences in the AFO

individuals showed the atypical patterns at this time point. Note that only some of the AFO individuals showed the atypical patterns, not all of them. The mystery of why such alterations occur only in a few individuals will be solved in future studies. The vocal behavior of the mouse pups born to the aged father may model the developmental disturbance in human infant crying, which is altered by the aging of the father. Our study highlights the significance and probability of indicating subtle “atypical” or “neuro-diverse” behaviors in infancy for the diagnosis of ASD and related neurodevelopmental disorders.

Finally, we would like to emphasize that the study described here has the potential to lead to a quantitative analysis of individuality. Individual differences in behavior are a universal phenomenon in nature, but conventional analyses have mainly focused on “average” differences, and in a sense, ignored individual differences. To date, emerging efforts have been made in the field of neuroscience to better understand individuality. In some studies of behavioral and brain development, individual differences have been assessed as variables of interest rather than treated as statistical noise (Foulkes and Blakemore, 2018; Kidd et al., 2018; Becht and Mills, 2020). Although our VAE lacks comparison analyses with a large number of literature and other advanced approaches such as “Nearest Neighbours” (Kollmorgen et al., 2020), our study paves the way for applying a novel non-supervised machine-learning approach to the re-analysis of existing databases of other animals and human infant crying. In the future, the VAE has incredible potential for discoveries of translatable key biomarkers and

analyses of the individual difference between animals and humans, leading to the quantitative research of individuality.

Author Contributions

Designed research: L. M. and N. O.

Developed paternal aging model, collected the USV of the paternal aging model, and performed the USV analyses of the two models: L. M.

Contributed analytic tools: K. K., R. O. T

Analyzed USV data: L. M., R. K., H. I., T. M., R. O. T., F. K.

Developed *Auts2*$^{del8/+}$ model and collected the USV of *Auts2*$^{del8/+}$ model: K. H. and M.H.

Technique for USV recording: R.K., Y.K.

Wrote the paper: L. M., H. I., R. K., V.T., N. O.

Discussed and interpreted the results: all authors

Data availability

All relevant data are available from the authors upon request.

References

Acebo C, Thoman EB (1992) Crying as social behavior. Infant Mental Health Journal 13:67-82.

Acebo C, Thoman EB (1995) Role of infant crying in the early mother-infant dialogue. Physiology and Behavior 57:541-547.

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.). Washington, DC: American Psychiatric Association.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23:185-188.

Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, Chawarska K, Klin A, Ercan-Sencicek AG, Stillman AA, Tanriover G, Abrahams BS, Duvall JA, Robbins EM, Geschwind DH, Biederer T, Gunel M, Lifton RP, State MW (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet 82:165-173.

Becht AI, Mills KL (2020) Modeling Individual Differences in Brain Development. Biol Psychiatry 88:63-69.

Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D, Bonin M, Riess A, Engels H, Sprengel R, Scherer SW, Rappold GA (2010) Mutations in the SHANK2 synaptic scaffolding

gene in autism spectrum disorder and mental retardation. Nat Genet 42:489-491.

Branchi I, Santucci D, Alleva E (2001) Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. Behav Brain Res 125:49-56.

D'Amato F, Scalera E, Sarli C, Moles A (2005 Jan) Pups call, mothers rush: does maternal responsiveness affect the amount of ultrasonic vocalizations in mouse pups? . Behav Genet 35 (1):103-112.

Dougherty JD, Maloney SE, Wozniak DF, Rieger MA, Sonnenblick L, Coppola G, Mahieu NG, Zhang J, Cai J, Patti GJ, Abrahams BS, Geschwind DH, Heintz N (2013) The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. J Neurosci 33:2732-2753.

Dudek J, Faress A, Bornstein MH, Haley DW (2016) Infant cries rattle adult cognition. PLoS One 11:e0154283.

Esposito G, Hiroi N, Scattoni ML (2017) Cry, baby, cry: expression of distress as a biomarker and modulator in autism spectrum disorder. International Journal of Neuropsychopharmacology 20:498-503.

Foulkes L, Blakemore SJ (2018) Studying individual differences in human adolescent brain development. Nat Neurosci 21:315-323.

Gerlach M, Farb B, Revelle W, Nunes Amaral LA (2018) A robust data-driven approach identifies four personality types across four large data sets. Nat Hum Behav 2:735-742.

Goffinet J, Mooney R, Pearson J (2019) Inferring low-dimensional latent descriptions of animal vocalizations. bioRxiv.

Goldberg LR (1990) An alternative "description of personality": the big-five factor structure. J Pers Soc Psychol 59:1216-1229.

Grimsley JM, Monaghan JJ, Wenstrup JJ (2011) Development of social vocalizations in mice. PLoS One 6:e17460.

Grzadzinski R, Huerta M, Lord C (2013) DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. Mol Autism 4.

Hahn ME, Lavooy MJ (2005) A Review of the methods of studies on infant ultrasound production and maternal retrieval in small rodents. Behav Genet 35 (1):31-52.

Homberg JR et al. (2016) Understanding autism and other neurodevelopmental disorders through experimental translational neurobehavioral models. Neurosci Biobehav Rev 65:292-312.

Hori K et al. (2020) AUTS2 Regulation of Synapses for Proper Synaptic Inputs and Social Communication. iScience 23:101183.

Hultman CM, Sandin S, Levine SZ, Lichtenstein P, Reichenberg A (2011) Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. Mol Psychiatry 16:1203-1212.

Janecka M, Manduca A, Servadio M, Trezza V, Smith R, Mill J, Schalkwyk LC, Reichenberg A, Fernandes C (2015) Effects of advanced paternal age on trajectories of social behavior in offspring. Genes Brain Behav 14:443-453.

Jarvis ED (2019) Evolution of vocal learning and spoken language. Science 366:50-54.

Kidd E, Donnelly S, Christiansen MH (2018) Individual Differences in Language Acquisition and Processing. Trends Cogn Sci 22:154-169.

Kikusui T, Nakanishi K, Nakagawa R, Nagasawa M, Mogi K, Okanoya K (2011) Cross fostering experiments suggest that mice songs are innate. PLoS ONE 6 (3):e17721.

Kojima M, Yassin W, Owada K, Aoki Y, Kuwabara H, Natsubori T, Iwashiro N, Gonoi W, Takao H, Kasai K, Abe O, Kano Y, Yamasue H (2019) Neuroanatomical Correlates of Advanced Paternal and Maternal Age at Birth in Autism Spectrum Disorder. Cereb Cortex 29:2524-2532.

Kollmorgen S, Hahnloser RHR, Mante V (2020) Nearest neighbours reveal fast and slow components of motor learning. Nature 577:526-530.

Konishi S, Kitagawa G (2008) Information Criteria and Statistical Modeling.: Springer-Verlag New York Press.

Lai JK, Sobala-Drozdowski M, Zhou L, Doering LC, Faure PA, Foster JA (2014) Temporal and spectral differences in the ultrasonic vocalizations of fragile X knock out mice during postnatal development. Behav Brain Res 259:119-130.

Lopatina OL, Furuhara K, Ishihara K, Salmina AB, Higashida H (2017) Communication Impairment in Ultrasonic Vocal Repertoire during the Suckling Period of Cd157 Knockout Mice: Transient Improvement by Oxytocin. Frontiers in Neuroscience 11.

Loth E et al. (2017) The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. Mol Autism 8:24.

Lotspeich LJ, Ciaranello RD (1993) The neurobiology and genetics of infantile autism. Int Rev Neurobiol 35:87-129.

McLachlan G, Peel D (2000) Finite Mixture Models., 1 st Edition: Wiley-Interscience Press.

Mooney R (2020) The neurobiology of innate and learned vocalizations in rodents and songbirds. Current opinion in neurobiology 64:24-31.

Nardone S, Sams DS, Zito A, Reuveni E, Elliott E (2017) Dysregulation of cortical neuron DNA methylation profile in autism spectrum disorder. Cereb Cortex 27:5739-5754.

Nave G, Minxha J, Greenberg DM, Kosinski M, Stillwell D, Rentfrow J (2018) Musical Preferences Predict Personality: Evidence From Active Listening and Facebook Likes. Psychol Sci 29:1145-1158.

Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, Harlap S, Rabinowitz J, Shulman C, Malaspina D, Lubin G, Knobler HY, Davidson M, Susser E (2006) Advancing paternal age and autism. Arch Gen Psychiatry 63(9):1026-1032.

Romano E, Michetti C, Caruso A, Laviola G, Scattoni ML (2013) Characterization of neonatal vocal and motor repertoire of reelin mutant mice. PLoS One 8:e64407.

Sampino S, Juszczak GR, Zacchini F, Swiergiel AH, Modlinski JA, Loi P, Ptak GE (2014) Grand-paternal age and the development of autism-like symptoms in mice progeny. Transl Psychiatry 4:e386.

Sandin S et al. (2016) Autism risk associated with parental age and with increasing difference in age between the parents. Mol Psychiatry 21:693-700.

Scattoni ML, Crawley J, Ricceri L (2009) Ultrasonic vocalizations: a tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. Neurosci Biobehav Rev 33:508-515.

Scattoni ML, Gandhy SU, Ricceri L, Crawley JN (2008) Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. PLoS One 3:e3067.

Schaefer GB, Mendelsohn NJ (2008) Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. Genet Med 10:4-12.

Sheinkopf SJ, Iverson JM, Rinaldi ML, Lester BM (2012) Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. Autism Res 5:331-339.

Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, De Gasperi R, Sosa MA, Rabidou D, Santucci AC, Perl D, Morrisey E, Buxbaum JD (2005) Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A 102:9643-9648.

Soltis J (2004) The signal functions of early infant crying. Behav Brain Sci 27:443-490.

Sultana R, Yu CE, Yu J, Munson J, Chen D, Hua W, Estes A, Cortes F, de la Barra F, Yu D, Haider ST, Trask BJ, Green ED, Raskind WH, Disteche CM, Wijsman E, Dawson G, Storm DR, Schellenberg GD, Villacres EC (2002) Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. Genomics 80:129-134.

Szatmari P et al. (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39:319-328.

Tachibana RO, Kanno K, Okabe S, Kobayasi KI, Okanoya K (2020) USVSEG: A robust method for segmentation of ultrasonic vocalizations in rodents. PLoS One 15:e0228907.

Takahashi T, Okabe S, Broin PO, Nishi A, Ye K, Beckert MV, Izumi T, Machida A, Kang G, Abe S, Pena JL, Golden A, Kikusui T, Hiroi N (2016) Structure and function of neonatal social communication in a genetic mouse model of autism. Mol Psychiatry 21:1208-1214.

Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G, Bonnot O, Weismann-Arcache C, Botbol M, Lauth B, Ginchat V, Roubertoux P, Barburoth M, Kovess V, Geoffray M-M, Xavier J (2014) Gene x environment interactions in autism spectrum disorders: role of epigenetic mechanisms. Frontiers in Psychiatry 5.

Unwin LM, Bruz I, Maybery MT, Reynolds V, Ciccone N, Dissanayake C, Hickey M, Whitehouse AJO (2017) Acoustic properties of cries in 12-Month old

infants at high-risk of autism spectrum disorder. J Autism Dev Disord 47:2108-2119.

Wohr M (2014) Ultrasonic vocalizations in Shank mouse models for autism spectrum disorders: detailed spectrographic analyses and developmental profiles. Neurosci Biobehav Rev 43:199-212.

Wozniak RH, Leezenbaum NB, Northrup JB, West KL, Iverson JM (2017) The development of autism spectrum disorders: variability and causal complexity. Wiley Interdiscip Rev Cogn Sci 8:e1426.

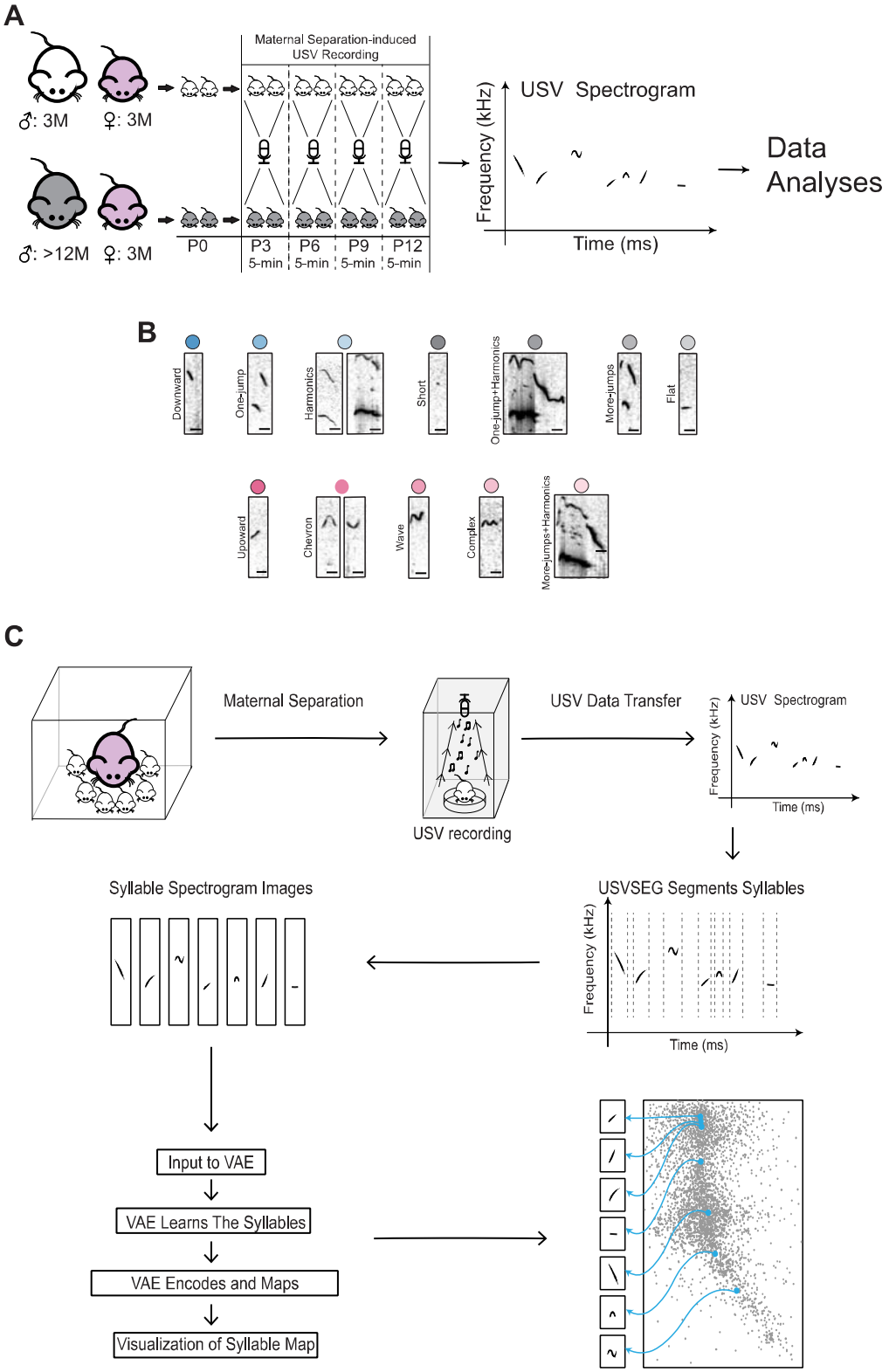
Yoshizaki K, Koike K, Kimura R, Osumi N (2017) Early postnatal vocalizations predict sociability and spatial memory in C57BL/6J mice: Individual differences in behavioral traits emerge early in development. PLoS One 12:e0186798.

Yoshizaki K, Furuse T, Kimura R, Tucci V, Kaneda H, Wakana S, Osumi N (2016) Paternal Aging Affects Behavior in Pax6 Mutant Mice: A Gene/Environment Interaction in Understanding Neurodevelopmental Disorders. PLoS One 11:e0166665.

Yoshizaki K, Kimura R, Kobayashi H, Oki S, Kikkawa T, Mai L, Koike K, Mochizuki K, Inada H, Matsui Y, Kono T, Osumi N (2021) Paternal age affects offspring via an epigenetic mechanism involving REST/NRSF. EMBO reports:e51524.

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley JC, Warren ST, Schlessinger D (1991) Fragile X genotype characterized by an unstable region of DNA. Science 252:1179-1181.

Figure 1


A
Maternal Separation-induced USV Recording
♂: 3M
♀: 3M
♂: >12M
♀: 3M
P0
P3 5-min
P6 5-min
P9 5-min
P12 5-min
USV Spectrogram
Frequency (kHz)
Time (ms)
Data Analyses
B
Downward
One-jump
Harmonics
Short
One-jump+Harmonics
More-jumps
Flat
Upward
Chevron
Wave
Complex
More-jumps+Harmonics
C
Maternal Separation
USV recording
USV Data Transfer
USV Spectrogram
Frequency (kHz)
Time (ms)
USVSEG Segments Syllables
Syllable Spectrogram Images
Input to VAE
VAE Learns The Syllables
VAE Encodes and Maps
Visualization of Syllable Map

735

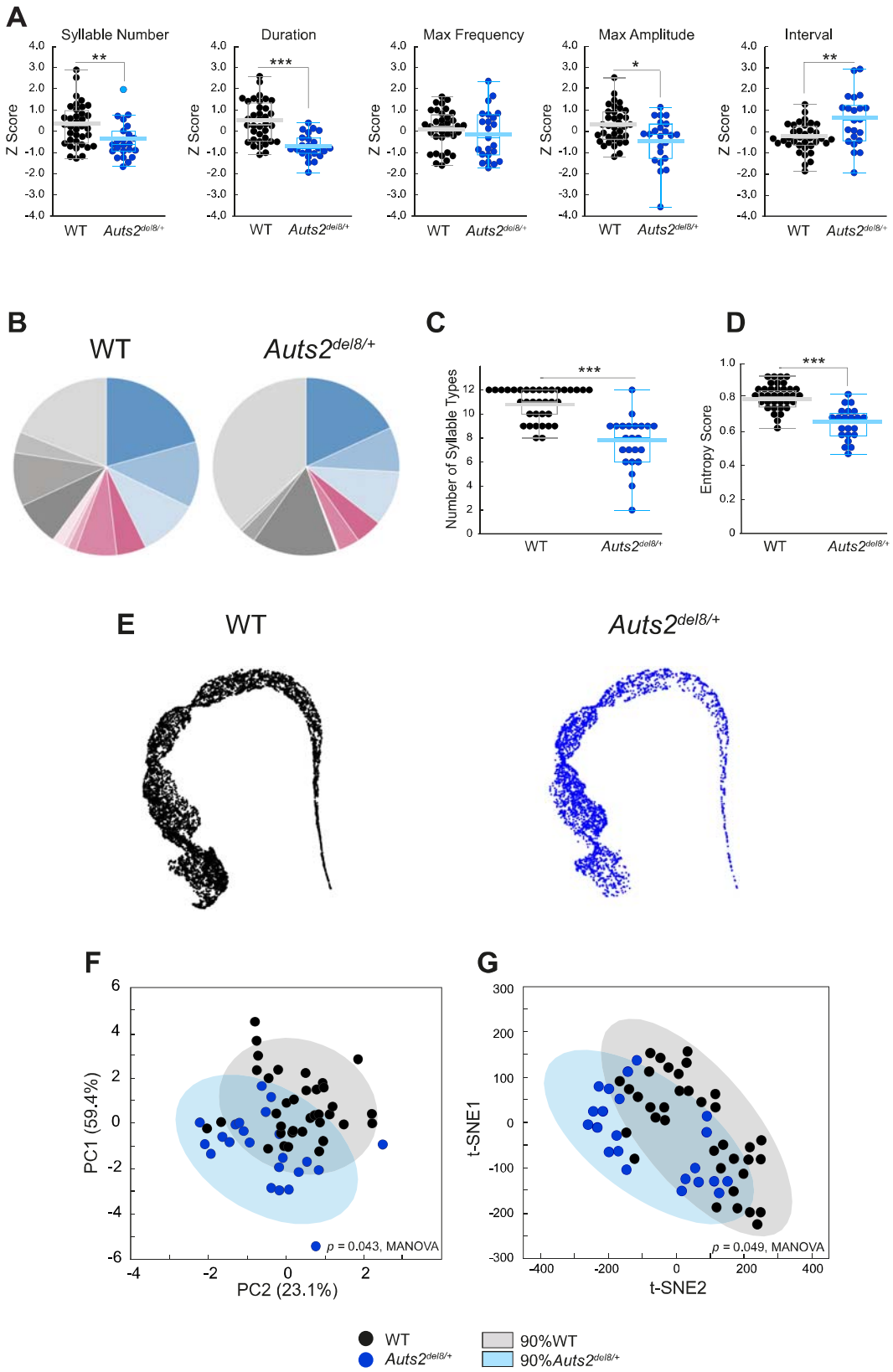
Figure 1. Schematic diagram of the Experimental design and USV analyses.

(A) The experimental design in the model of paternal aging.

(B) Typical sonograms of twelve types of syllables. Scale bar = 60 kHz, 20 ms.

(C) Following maternal separation, USVs were recorded from individual pups one by one. The resulting USVs were processed to spectrogram by USVSEG, which segmented the syllables and outputted the syllable images with acoustic parameters. Subsequently, the VAE was trained to learn and encode the syllables. The resulting syllables were visualized and mapped in a latent space.

Figure 2


A
Syllable Number
Duration
Max Frequency
Max Amplitude
Interval
Z Score
WT
Auts2del8/+
B
WT
Auts2del8/+
C
Number of Syllable Types
D
Entropy Score
E
WT
Auts2del8/+
F
PC1 (59.4%)
PC2 (23.1%)
p = 0.043, MANOVA
G
t-SNE1
t-SNE2
p = 0.049, MANOVA
WT
Auts2del8/+
90%WT
90%Auts2del8/+

744

Figure 2. *Auts2*$^{del8/+}$ pups display the USV impairments at P7.

(A) Z scores of syllable number, duration, maximum frequency, maximum amplitude, and interval. The thick horizontal lines indicate the mean value of Z scores. Boxplots represent the range of Z scores. The horizontal lines in the boxes signify respective group medians, boxes are 25th – 75th percentiles, and whiskers represent minimum to maximum. The light blue point depicts the outlier. Each point represents the syllable data of one pup. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ indicate a significant difference between the wildtype and *Auts2*$^{del8/+}$ pups (Wilcoxon test).

(B) Pie graphs show the percentage composition of the twelve types of syllables in the pups of wildtype and *Auts2*$^{del8/+}$. The colors indicate the syllable type shown in Figure 1B.

(C) A significant decrease of the number of syllable types in *Auts2*$^{del8/+}$ pups. The thick horizontal lines indicate the mean value of the number of syllable types. Boxplots represent the range of the number of syllable types. The horizontal lines in the boxes signify respective group medians, boxes are 25th – 75th percentiles, and whiskers represent minimum to maximum. Each point represents the data of one pup. *** $p < 0.001$ indicates a significant difference (Wilcoxon test).

(D) A significant decrease of entropy scores of syllables in *Auts2*$^{del8/+}$ pups. The thick horizontal lines indicate the mean value of entropy scores. Boxplots represent the range of entropy scores. The horizontal lines in the boxes signify respective group medians, boxes are 25th – 75th percentiles, and whiskers

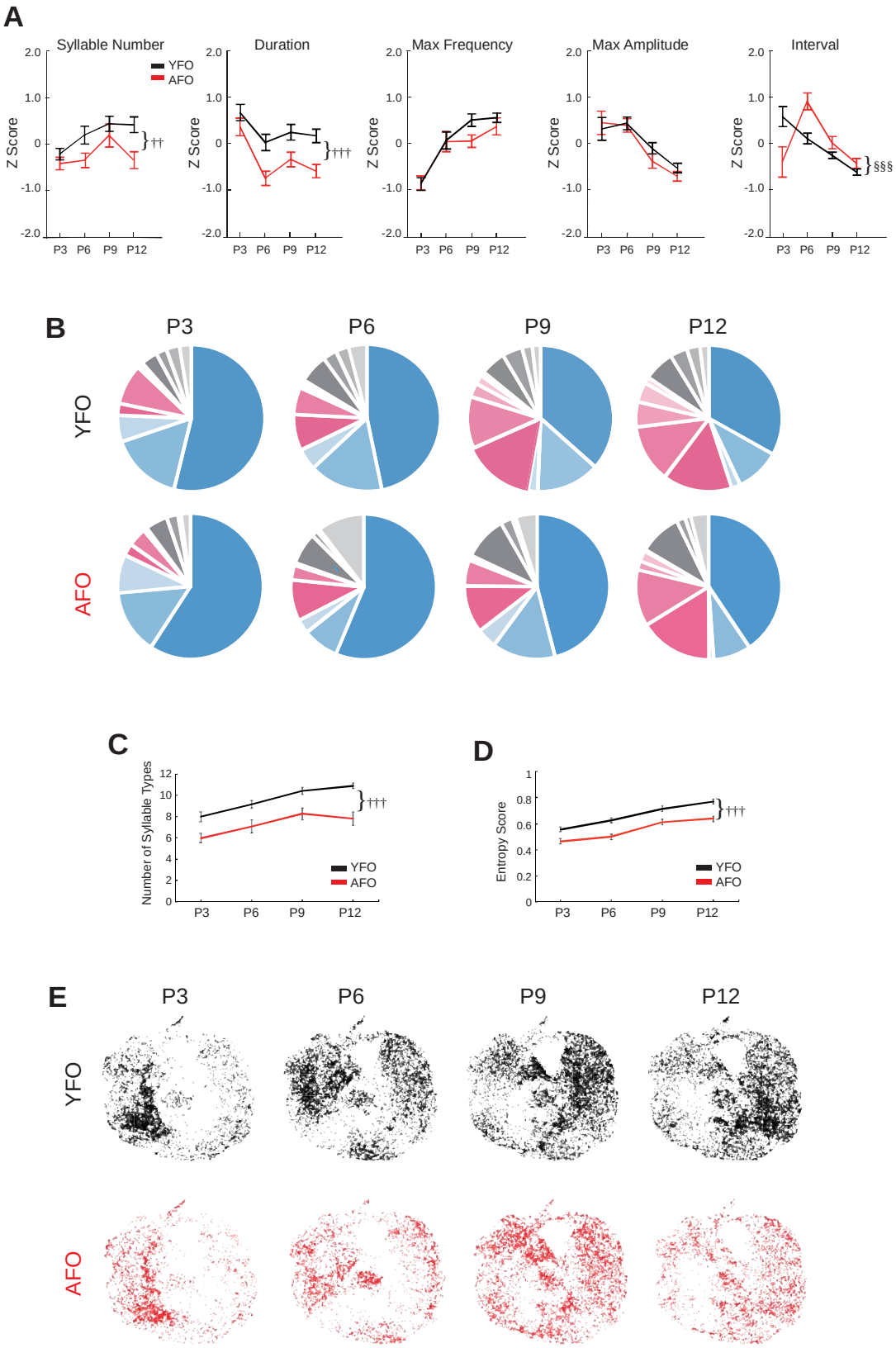
represent minimum to maximum. Each point represents the data of one pup. *** $p < 0.001$ indicates a significant difference (Wilcoxon test).

(E) The VAE takes spectrograms from the syllables of wildtype and *Auts2*$^{del8/+}$ pups as input and visualizes them as a vector of latent dimensions at P7. Each point in this latent space represents a single spectrogram of a syllable. The distance between points indicates neighbor similarity.

(F) Based on the USVSEG, PCA summarized the patterns of individual wildtype and *Auts2*$^{del8/+}$ pups. Each point represents the syllable pattern of one pup. Shorter distances between points indicate higher similarity in syllable patterns. Grey and light blue clusters contain 90% of the population of wildtype and *Auts2*$^{del8/+}$ pups, respectively.

(G) Based on the VAE, t-SNE summarized the patterns of individual wildtype and *Auts2*$^{del8/+}$ pups. Each point represents the syllable patterns of one pup. Shorter distances between points indicate higher similarity in syllable patterns. Grey and light blue clusters contain 90% of the population of wildtype and *Auts2*$^{del8/+}$ pups, respectively.

Figure 3


A
Syllable Number
Duration
Max Frequency
Max Amplitude
Interval
YFO
AFO
Z Score
2.0
1.0
0
-1.0
-2.0
P3
P6
P9
P12
††
†††
§§§
B
P3
P6
P9
P12
YFO
AFO
C
Number of Syllable Types
12
10
8
6
4
2
0
YFO
AFO
D
Entropy Score
1
0.8
0.6
0.4
0.2
0
E
P3
P6
P9
P12
YFO
AFO

783

Figure 3. Paternal aging alters the syllable development in offspring.

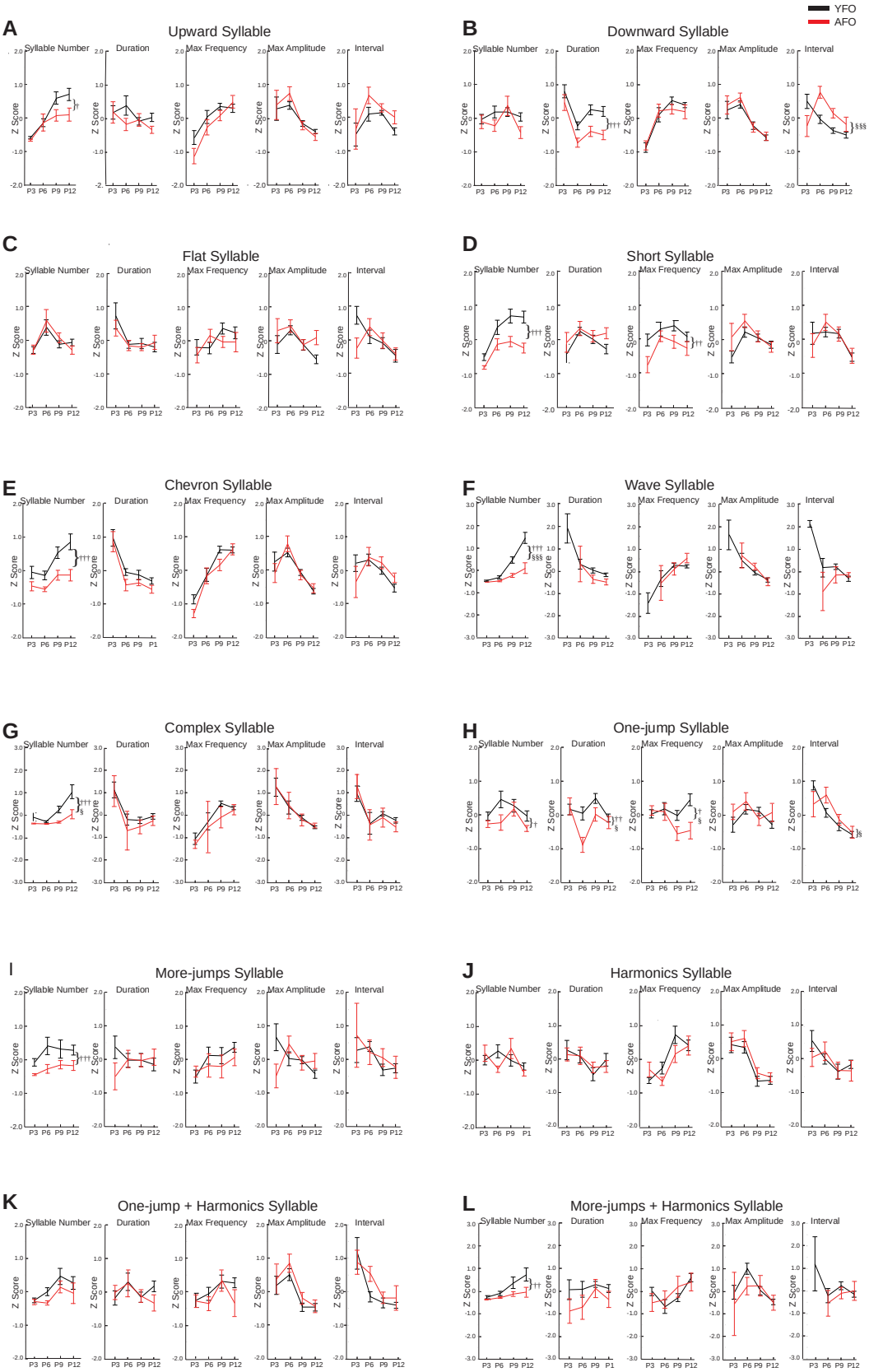
(A**)** Z scores of syllable number, duration, maximum frequency, maximum amplitude, and interval. †† $p < 0.01$, ††† $p < 0.001$ indicate a significant main effect of father's age; §§§ $p < 0.001$ indicates a significant interaction of father's age × postnatal day (ANOVA).

(B) Pie graphs show the percentage composition of the twelve types of syllables in the YFO and AFO from P3 to P12. The colors indicate the syllable type shown in Figure 1B. Magenta, blue, and grey syllable types indicate those that showed percentage increases, decreases, and no specific change in the YFO, respectively. (C) Developmental trajectories of the number of syllable types. ††† $p < 0.001$ indicates a significant main effect of the father's age (ANOVA).

(D) Normalized entropy scores of syllables. ††† $p < 0.001$ indicates a significant main effect of the father's age (ANOVA).

(E) The VAE takes spectrograms from the syllables of the YFO and AFO as input and visualizes them as a vector of latent dimensions from P3 to P12. Each point in this latent space represents a single spectrogram of a syllable. The distance between points indicates neighbor similarity.

Figure 4

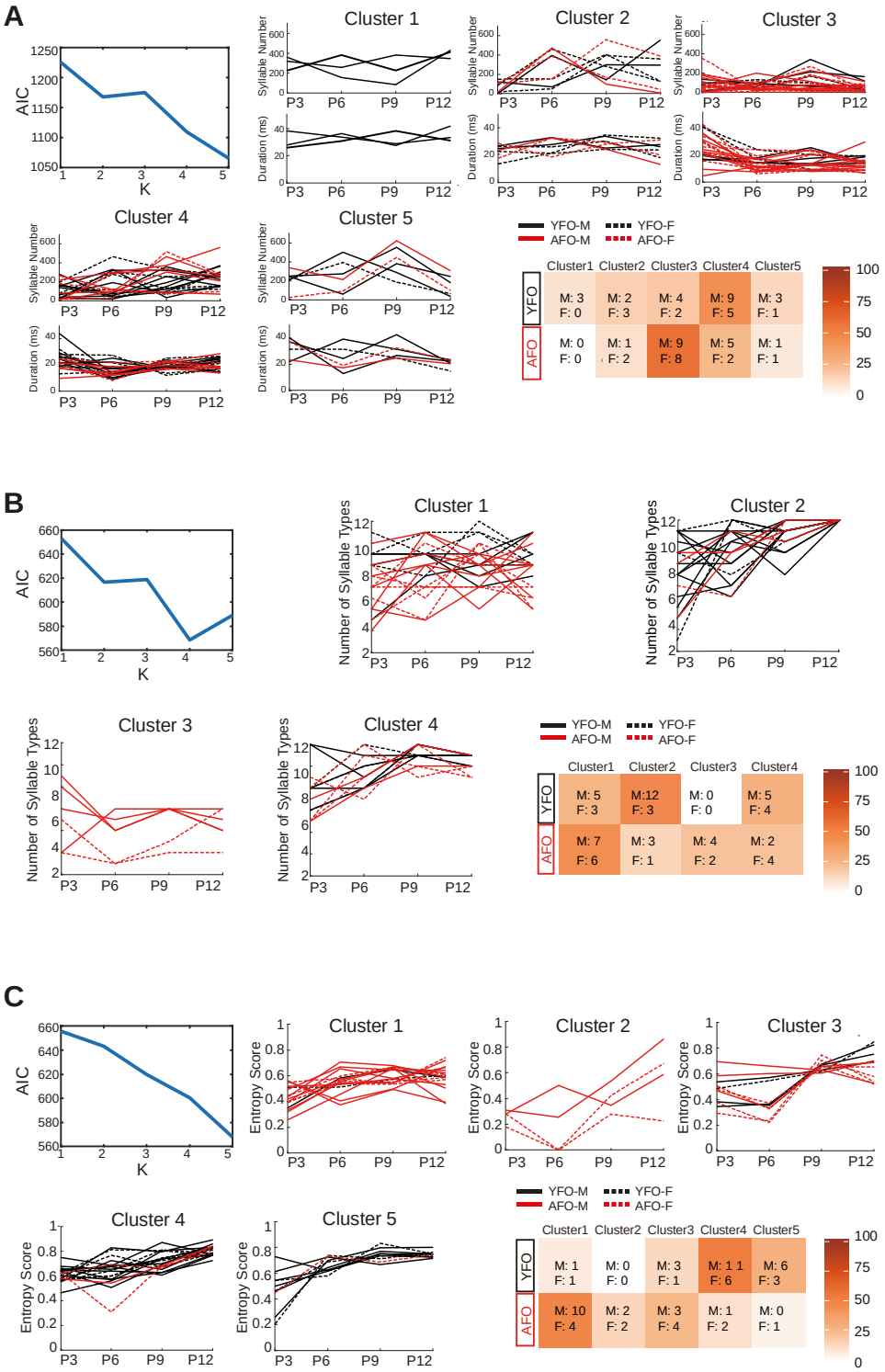

YFO
AFO
A
Upward Syllable
B
Downward Syllable
C
Flat Syllable
D
Short Syllable
E
Chevron Syllable
F
Wave Syllable
G
Complex Syllable
H
One-jump Syllable
I
More-jumps Syllable
J
Harmonics Syllable
K
One-jump + Harmonics Syllable
L
More-jumps + Harmonics Syllable
Syllable Number
Duration
Max Frequency
Max Amplitude
Interval
Z Score
P3 P6 P9 P12

801

Figure 4. Paternal aging alters the developmental trajectory of distinct types of syllables.

(A)-(L) show the developmental trajectories of the syllable number, duration, maximum frequency, maximum amplitude, and interval of distinct types of syllables from the YFO and AFO. † $p < 0.05$, †† $p < 0.01$, ††† $p < 0.001$ indicates a significant main effect of father's age. § $p < 0.05$, §§ $p < 0.01$, §§§ $p < 0.001$ indicates a significant interaction of father's age × postnatal day (ANOVA).

Figure 5


A
AIC
K
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Syllable Number
Duration (ms)
P3 P6 P9 P12
YFO-M
AFO-M
YFO-F
AFO-F
Cluster1 Cluster2 Cluster3 Cluster4 Cluster5
YFO
AFO
M: 3 F: 0 | M: 2 F: 3 | M: 4 F: 2 | M: 9 F: 5 | M: 3 F: 1
M: 0 F: 0 | M: 1 F: 2 | M: 9 F: 8 | M: 5 F: 2 | M: 1 F: 1
B
Number of Syllable Types
Cluster1 Cluster2 Cluster3 Cluster4
M: 5 F: 3 | M:12 F: 3 | M: 0 F: 0 | M: 5 F: 4
M: 7 F: 6 | M: 3 F: 1 | M: 4 F: 2 | M: 2 F: 4
C
Entropy Score
Cluster1 Cluster2 Cluster3 Cluster4 Cluster5
M: 1 F: 1 | M: 0 F: 0 | M: 3 F: 1 | M: 1 1 F: 6 | M: 6 F: 3
M: 10 F: 4 | M: 2 F: 2 | M: 3 F: 4 | M: 1 F: 2 | M: 0 F: 1

809

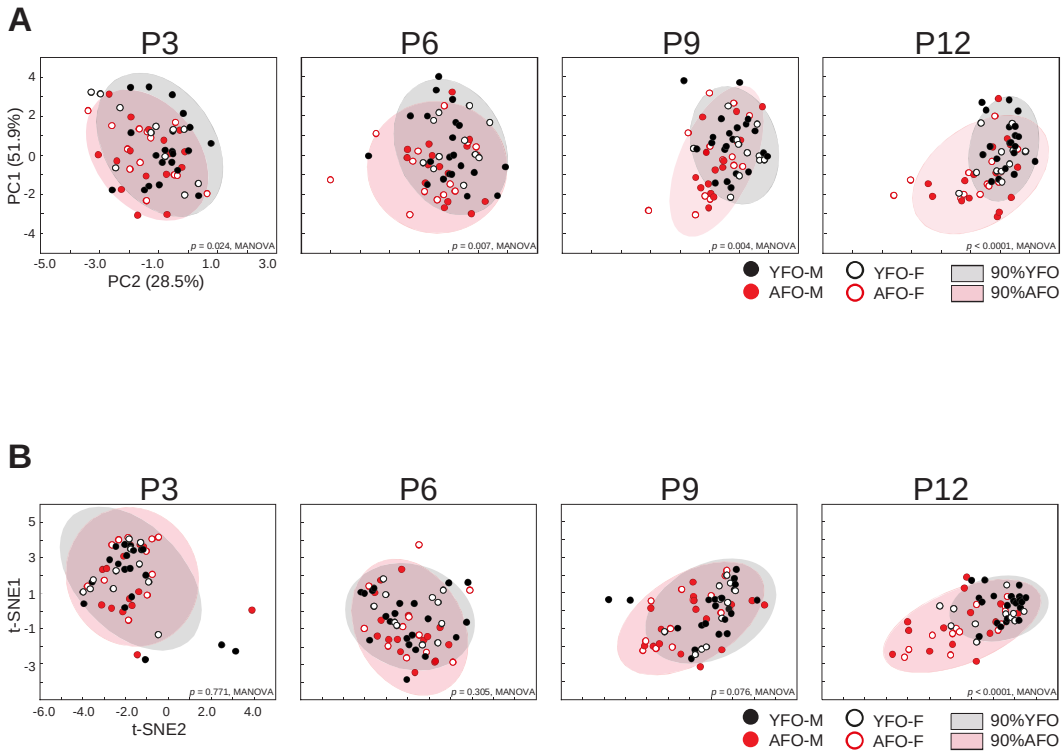
Figure 5. **The individual differences in USV longitudinal development.**

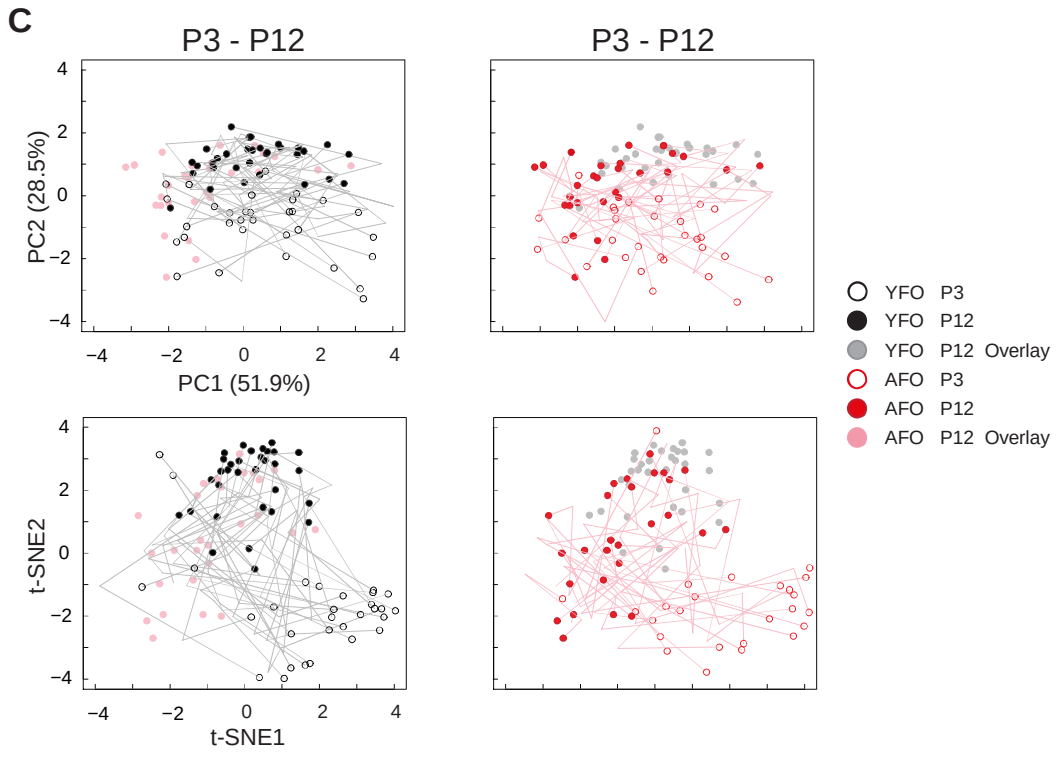
(A) Clustering analysis with GMMs separates the longitudinal developmental patterns among individuals according to syllable number and duration. Five was chosen as the number of clusters (shown on the X-axis) with the minimum AIC value (Y-axis). Clustering analyses separated individual offspring into five clusters. The colors of the heat map indicate the percentage of the YFO and AFO in each cluster. The numbers in the heat map indicate the numbers of pups. The M and F mean male and female, respectively.

(B) Clustering analysis with GMMs separates the longitudinal developmental patterns according to the number of syllable types. Four was chosen as the number of clusters (shown on the X-axis) with the minimum AIC value (Y-axis). Clustering analyses separated individual offspring into four clusters. The colors of the heat map indicate the percentage of the YFO and AFO in each cluster. The numbers in the heat map indicate the numbers of pups. The M and F mean male and female, respectively.

(C) Clustering analysis with GMMs separates the longitudinal developmental patterns according to the normalized entropy. Five was chosen as the number of clusters (shown on the X-axis) with the minimum AIC value (Y-axis). Clustering analyses separated individual offspring into five clusters. The colors of the heat map indicate the percentage of the YFO and AFO in each cluster. The numbers in the heat map indicate the numbers of pups. The M and F mean male and female, respectively.

Figure 6


A
P3
P6
P9
P12
PC1 (51.9%)
PC2 (28.5%)
p = 0.024, MANOVA
p = 0.007, MANOVA
p = 0.004, MANOVA
p < 0.0001, MANOVA
YFO-M
AFO-M
YFO-F
AFO-F
90%YFO
90%AFO
B
t-SNE1
t-SNE2
p = 0.771, MANOVA
p = 0.305, MANOVA
p = 0.076, MANOVA
p < 0.0001, MANOVA


C
P3 - P12
P3 - P12
PC2 (28.5%)
PC1 (51.9%)
t-SNE2
t-SNE1
YFO P3
YFO P12
YFO P12 Overlay
AFO P3
AFO P12
AFO P12 Overlay

832

Figure 6. The individual differences in USV transverse development.

(A) Based on the USVSEG, PCA summarized the patterns of individual YFO and AFO at P3, P6, P9, and P12. Each point represents the syllable pattern of one pup. Shorter distances between points indicate higher similarity in syllable patterns. Grey and light red clusters contain 90% of the population of the YFO and AFO, respectively. The M and F mean male and female, respectively.

(B) Based on the VAE, t-SNE summarized the patterns of individual YFO and AFO at P3, P6, P9, and P12. Each point represents the syllable features of one pup. Shorter distances between points indicate higher similarity in vocal repertoire. Grey and light red clusters contain 90% of the population of the YFO and AFO, respectively. The M and F mean male and female, respectively.

(C) Summary of individual traces at P3 and P12 based on (A) and (B). Each point represents the syllable pattern of one pup.

Table 1. The number of detected syllables.

The numbers of syllables from pups were detected by the USVSEG and VAE.

Approach	Model	Postnatal Day				
		P3	P6	P7	P9	P12
USVSEG	YFO	4212	6041		7122	7039
	AFO	2990	3276		5454	3296
VAE	YFO	3990	5980		6880	6860
	AFO	2800	3300		5450	3420
USVSEG	WT			8903		
	Auts2$^{del8/+}$			3564		
VAE	WT			8740		
	Auts2$^{del8/+}$			3480		

USVSEG: USV segmentation software

VAE: variational autoencoder program

YFO: young father-derived offspring

AFO: aged father-derived offspring

WT: wildtype offspring

Auts2$^{del8/+}$: the offspring with deleted exon 8 of the *Auts2* gene

Table 2. The *Auts2* mutant affects the percentage composition of distinct syllables.

MANOVA revealed the significant differences in syllable percentage composition between the WT and *Auts2*$^{del8/+}$.

Postnatal Day	Results of MANOVA	Syllable Type	Results of *t*-test
P7	F (11, 47) = 4.483, $p < 0.001$	Flat	18.67 ± 1.55%: WT vs 37.31 ± 2.64%: *Auts2*$^{del8/+}$, $p < 0.001$
		Short	8.06 ± 1.05%: WT vs 14.82 ± 2.41%: *Auts2*$^{del8/+}$, $p = 0.023$
		Chevron	7.17 ± 0.85%: WT vs 3.69 ± 0.85%: *Auts2*$^{del8/+}$, $p = 0.010$
		Wave	1.44 ± 0.27%: WT vs 0.20 ± 0.10%: *Auts2*$^{del8/+}$, $p < 0.001$
		Complex	0.92 ± 0.26%: WT vs 0.15 ± 0.10%: *Auts2*$^{del8/+}$, $p = 0.016$
		One-jump	12.05 ± 1.47%: WT vs 8.16 ± 1.04%: *Auts2*$^{del8/+}$, $p = 0.047$
		More-jumps	3.79 ± 0.69%: WT vs 0.68 ± 0.20%: *Auts2*$^{del8/+}$, $p < 0.001$
		One-jump + harmonics	9.76 ± 1.26%: WT vs 2.56 ± 0.72%: *Auts2*$^{del8/+}$, $p < 0.001$
		More-jumps + harmonics	1.99 ± 0.49%: WT vs 0.07 ± 0.05%: *Auts2*$^{del8/+}$, $p < 0.001$

WT: wildtype offspring

Auts2$^{del8/+}$: the offspring with deleted exon 8 of the *Auts2* gene

Table 3. Paternal aging affects the percentage composition of distinct syllables.

MANOVA revealed significant differences in syllable percentage composition between the YFO and AFO at P3, P6, P9, and P12.

Postnatal Day	Results of MANOVA	Syllable Type	Results of *t*-test
P3	F (11, 49) = 1.65, p = 0.114		
P6	F (11, 49) = 2.92, p = 0.005	Chevron	6.01 ± 0.63%: YFO vs 3.32 ± 0.57%: AFO, p = 0.031
		One-jump	16.27 ± 2.01%: YFO vs 8.03 ± 2.01%: AFO, p = 0.031
		More-jumps	2.87 ± 0.74%: YFO vs 0.71 ± 0.28%: AFO, p = 0.033
		One-jump + harmonics	3.05 ± 0.5%: YFO vs 1.35 ± 0.34%: AFO, p = 0.031
P9	F (11, 49) = 5.37, p < 0.001	Downward	36.54 ± 1.49%: YFO vs 46.03 ± 2.34%: AFO, p = 0.003
		Chevron	11.46 ± 0.78%: YFO vs 5.77 ± 0.87%: AFO, p < 0.001
		Wave	3.08 ± 0.43%: YFO vs 0.62 ± 0.16%: AFO, p < 0.001
		Complex	2.08 ± 0.40%: YFO vs 0.29 ± 0.10%: AFO, p < 0.001
		More-jumps	2.22 ± 0.47%: YFO vs 0.87 ± 0.22%: AFO, p = 0.029
		More-jumps +	0.86 ± 0.20%: YFO vs 0.20 ±

		harmonics	0.07%: AFO, *p* = 0.011
P12	F (11, 49) = 2.73, *p* = 0.008	Downward	33.14 ± 1.53%: YFO vs 40.57 ± 2.52%: AFO, *p* = 0.037
		Wave	5.51 ± 0.71%: YFO vs 2.00 ± 0.47%: AFO, *p* = 0.002
		More-jumps	2.88 ± 0.40%: YFO vs 1.31 ± 0.47%: AFO, *p* = 0.037
		More-jumps + harmonics	1.45 ± 0.31%: YFO vs 0.31 ± 0.14%: AFO, *p* = 0.011

YFO: young father-derived offspring

AFO: aged father-derived offspring

858 **Extended Data**

Procedures of VAE analyses

1. The recorded vocalizations were transferred to an UltraSound Gate 416H detector set (Avisoft Bioacoustics, Germany) at 25–125 kHz and processed to acoustic waveforms.
2. The USVSEG, which is a MATLAB script, was downloaded from https://sites.google.com/view/rtachi/resources.
3. The files of the acoustic waveforms were processed by USVSEG to segment the syllables. The parameters of USVSEG were set as follows: sampling (kHz): 300; time step (ms): 0.5; frequency min (kHz): 60 for the model of paternal aging and 40 for the model of *Auts2* mutant; frequency max (kHz): 120 for the model of paternal aging and 160 for the model of *Auts2* mutant; threshold (SD): 2.5; duration min (ms): 2; duration max (ms): 300; gap min (ms): 10; margin (ms): 20.
4. After the automatic procedure of the program script, each syllable was detected, and its onset/offset time was exported as .txt file.
5. The Python package: Autoencoded Vocal Analysis (AVA), which was used to generate, plot, and analyze latent features, was downloaded online: https://github.com/jackgoffinet/autoencoded-vocal-analysis.
6. The files of the acoustic waveforms and the .txt file of syllable onset/offset time were fed into VAE and preprocessed through the ‘preprocessing step’ to determine the visualization images. This step can be kept tuning until the optimization of spectrograms was produced:

```
from ava.preprocessing.utils import get_spec # makes spectrograms
```

```
from ava.models.vae import X_SHAPE # spectrogram dimensions

preprocess_params = {
    'get_spec': get_spec, # spectrogram maker
    'max_dur': 0.2 , # maximum syllable duration
    'min_freq': 60000 for the model of paternal aging and 40000 for the model of
Auts2 mutant, # minimum frequency
    'max_freq': 125000 or the model of paternal aging and 160000 for the model of
Auts2 mutan, # maximum frequency
    'num_freq_bins': X_SHAPE[0], # hard-coded
    'num_time_bins': X_SHAPE[1], # hard-coded
    'nperseg': 1024, # FFT
    'noverlap': 512, # FFT
    'spec_min_val': 0.5 for the model of paternal aging and 1.2 for the model of
Auts2 mutant, # minimum log-spectrogram value
    'spec_max_val': 3.5 for the model of paternal aging and 6 for the model of
Auts2 mutant, # maximum log-spectrogram value
    'fs': 250000, # audio samplerate
    'mel': False, # frequency spacing, mel or linear
    'time_stretch': True, # stretch short syllables?
    'within_syll_normalize': False, # normalize spectrogram values on a
                                    # spectrogram-by-spectrogram basis
    'max_num_syllables': None, # maximum number of syllables per directory
    'sylls_per_file': 10, # syllable per file
```

```
    'real_preprocess_params': ('min_freq', 'max_freq', 'spec_min_val', \
            'spec_max_val', 'max_dur'), # tunable parameters
    'int_preprocess_params': ('nperseg','noverlap'), # tunable parameters
    'binary_preprocess_params': ('time_stretch', 'mel', \
            'within_syll_normalize'), # tunable parameters
}

from ava.preprocessing.preprocess import tune_syll_preprocessing_params
audio_dirs = [...] # directories containing audio
seg_dirs = [...] # directories containing onset/offset decisions
preprocess_params = tune_syll_preprocessing_params(audio_dirs, seg_dirs, \
                preprocess_params)
```

7. The VAE was trained to maximize the ELB as the standard-setting. The latent dimension was fixed to 32. The inputted segmented spectrograms were split to a ratio of 80:20 randomly. Consequently, 80% and 20% of the spectrograms were used for a training set and a test set, respectively:

```
spec_dirs = [...] # directories containing saved spectrograms (hdf5s)
split = 0.8 # 80/20 train/test split

# Construct a random train/test partition.
from ava.models.vae_dataset import get_syllable_partition
partition = get_syllable_partition(spec_dirs, split)
```

```
# Make Dataloaders.
from ava.models.vae_dataset import get_syllable_data_loaders
loaders = get_syllable_data_loaders(partition)
```

After setting, the training was started as follows:

```
# Construct network.
from ava.models.vae import VAE
save_dir = 'model/parameters/should/be/saved/here/'
model = VAE(save_dir=save_dir)

# Train.
model.train_loop(loaders, epochs=101
```

Following the first training, a reconstruction of the detected syllable spectrogram was displayed as a latent map. In this map, each point represents a single spectrogram image. The cluster of the false-positive syllables on the latent map was identified and removed manually as background noise.

8. The VAE was trained again by using the saved model after the background noise deletion:

```
# Make an untrained model.
model = VAE(save_dir=save_dir)
```

```

# Load saved state.
model_checkpoint = 'path/to/checkpoint_100.tar'
model.load_state(model_checkpoint)

# Train another few epochs.
model.train_loop(loaders, epochs=101
```

9. The DataContainer object of AVA was used to bind all the data to manage the large data files and simplify the data analysis. The pups emitted less than ten calls were excluded from this analysis because data loader creation requires a minimum of 10 syllables.

```
import os
from ava.data.data_container import DataContainer

root = 'path/to/project/'
animals = ['animal_1', 'animal_2']
audio_dirs = [os.path.join(root, animal, 'audio') for animal in animals]
feature_dirs = [os.path.join(root, animal, 'features') for animal in animals]
spec_dirs = [os.path.join(root, animal, 'specs') for animal in animals]
projection_dirs = [os.path.join(root, animal, 'projs') for animal in animals]
model_fn = os.path.join(root, 'checkpoint_100.tar')

```

```
dc = DataContainer(audio_dirs=audio_dirs, feature_dirs=feature_dirs, \
            spec_dirs=spec_dirs, projection_dirs=projection_dirs, \
            model_filename=model_fn)
```

10. The data was finally analyzed and visualized as follows:

```
from ava.plotting.latent_projection import latent_projection_plot_DC

dc = ... # define DataContainer
latent_projection_plot_DC(dc, embedding_type='latent_mean_umap')
```