Analyses of infant vocal communication in mouse models of autism spectrum disorder using a novel unsupervised machine learning approach

(117/120 characters)

Lingling Mai[1], Ryuichi Kimura[2], Kei Hori[3], Hitoshi Inada[1,4], Kouta Kanno[5], Takeru Matsuda[6], Ryosuke O. Tachibana[7], Kaichi Yoshizaki[8], Valter Tucci[9], Fumiyasu Komaki[10,11], Mikio Hoshino[3], Noriko Osumi[1]*

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

[2]Department of Drug Discovery Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, 606-8501.

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

[4]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.

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

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

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

[8]Department of Disease Model, Institute for Developmental Research, Aichi Developmental Disability Center, Aichi, Japan, 480-0392.

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

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

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

* Corresponding author

This PDF file includes:

Main Text
Figures 1 to 6
Tables 1
Supplementary Table 1 to 2

Abstract (136/150 words)
Infants at a high risk of autism spectrum disorder show atypical patterns of crying, i.e., an innate behavior of social communication, which are modeled in ultrasonic vocalization (USV) in mouse pups. Here we present a novel unsupervised machine learning approach (a variational autoencoder, VAE) that can capture comprehensive features from a simple recording of USV. We studied offspring born to aged fathers and those derived from *Autism susceptibility candidate 2 (Auts2)* gene mutant fathers. These model mouse pups show fewer USVs, altered acoustic features and narrower diversity, highlighting overall atypical patterns of USV with a wide variability of phenotypes. The two models also reveal different USV distribution patterns in an inferring map. The VAE approach reveals potential merit in understanding complex phenotypes in individual vocal communication based on big data obtained automatically within a short period.

Introduction

Infant crying is an innate behavior of social communication (Acebo & Thoman, 1992; Acebo & Thoman, 1995) that is used to attract attention from caregivers (Soltis, 2004) and affects 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 a sign of a high 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: impairment of social interaction (including verbal and nonverbal communication deficits) and repetitive behavior (American Psychiatric Association, 2013; Lotspeich & Ciaranello, 1993). Although the etiology of ASD remains unclear, environmental, genetic, and epigenetic factors have been suggested to be intertwined (Nardone et al., 2017; Schaefer & Mendelsohn, 2008; Tordjman et al., 2014). For instance, recent epidemiological studies have suggested a significant association between paternal aging and the risk of ASD in offspring (Hultman et al., 2011; Reichenberg et al., 2006). On the other hand, 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), and SHANK2*(Berkel et al., 2010). Recently, the *Autism susceptibility candidate 2 (AUTS2)* gene mutations were identified in individuals with ASD (Bakkaloglu et al., 2008; Sultana et al., 2002). Thus, it is pivotal to examine whether infant crying is affected by these risk factors, and if so, whether they promote early developmental deficiencies linked to ASD, and whether similar and different phenotypes exist between potentially distinct etiologies.

Vocal communication in animals has been extensively studied in songbirds. In contrast, ultrasonic vocalization (USV) in rodents, especially in mice, has 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 can trigger maternal approach and retrieval behavior (D'Amato et al., 2005 Jan; Hahn & Lavooy, 2005). Thus, maternal separation-induced USV in mice is considered a form of social communication that corresponds to human infant crying (Esposito et al., 2017). During normal postnatal development, USV “syllables” from pups undergo gradual changes in acoustic features and components (Grimsley et al., 2011; Scattoni et al., 2008). Previously, ASD model mice have been shown to exhibit differences in various USV parameters, including being fewer in number, having higher or lower syllable frequency, and shorter syllable duration (Dougherty et al., 2013; Scattoni et al., 2008; Shu et al., 2005; Wohr, 2014). However, USV analyses commonly based on some hand-picked features that could lose some important information.

In this study, we applied a novel unsupervised modeling approach to comprehensively understand the genetic and non-genetic effects on maternal separation-induced USV in neonatal mice. We demonstrated that developmental differences of USVs obtained in the two different models of paternal aging (i.e., a

non-genetic model) and *Auts2* mutation (i.e., a genetic model), paying careful attention to individual variability in syllable patterns. Based on the results, we conclude and confirm that though the two models of paternal aging and of *Auts2* mutant show distinct syllable maps, both of them alter quantitatively and qualitatively the syllable development, then lead to an increase in the number of offspring with "atypical" USV patterns during early infancy. Thus, it reflects similar phenotypes that is present in the infants with atypical crying, and this phenomenon might be induced by either non-genetic or genetic etiology.

Results

Paternal aging alters the acoustic features of syllables emitted by offspring

We applied a semi-automatic procedure using USV segmentation software (USVSEG) and an unsupervised modeling approach using a variational autoencoder (VAE) to analyze USV syllables (see detailed procedures in Methods, **Fig.1** and **Appendix**). Offspring were obtained by mating young wild type female mice with either young (3 months) or aged (>12 months) wild type male mice (YFO and AFO, respectively, **Fig.2A**). A large number of syllables were collected from YFO and AFO at postnatal day 3 (P3), P6, P9, and P12 (**Table 1**). USVSEG output data showed that paternal aging altered the trajectories of overall syllables emitted by AFO, as reflected by a reduced number and shorter duration of syllables, and different intervals between adjacent syllables (**Fig.2B**).

To obtain a deeper understanding, we first took a conventional approach by categorizing all syllables into twelve types (**Fig. 2C**) based on their spectrotemporal patterns, according to previous reports (Hori et al., 2020; Kikusui et al., 2011; Scattoni et al., 2008). The USV acoustic features of the twelve syllable types are shown in **Fig. 3**. 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 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. 2D** and Supplementary **Table 1**). While no statistical difference was detected at P3, AFO emitted USVs with significantly different syllable components to those of YFO from P6 to P12. Compared to the syllables from YFO on each postnatal day, 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 AFO during postnatal development. These results show that paternal aging led to significantly different compositions of distinct syllables.

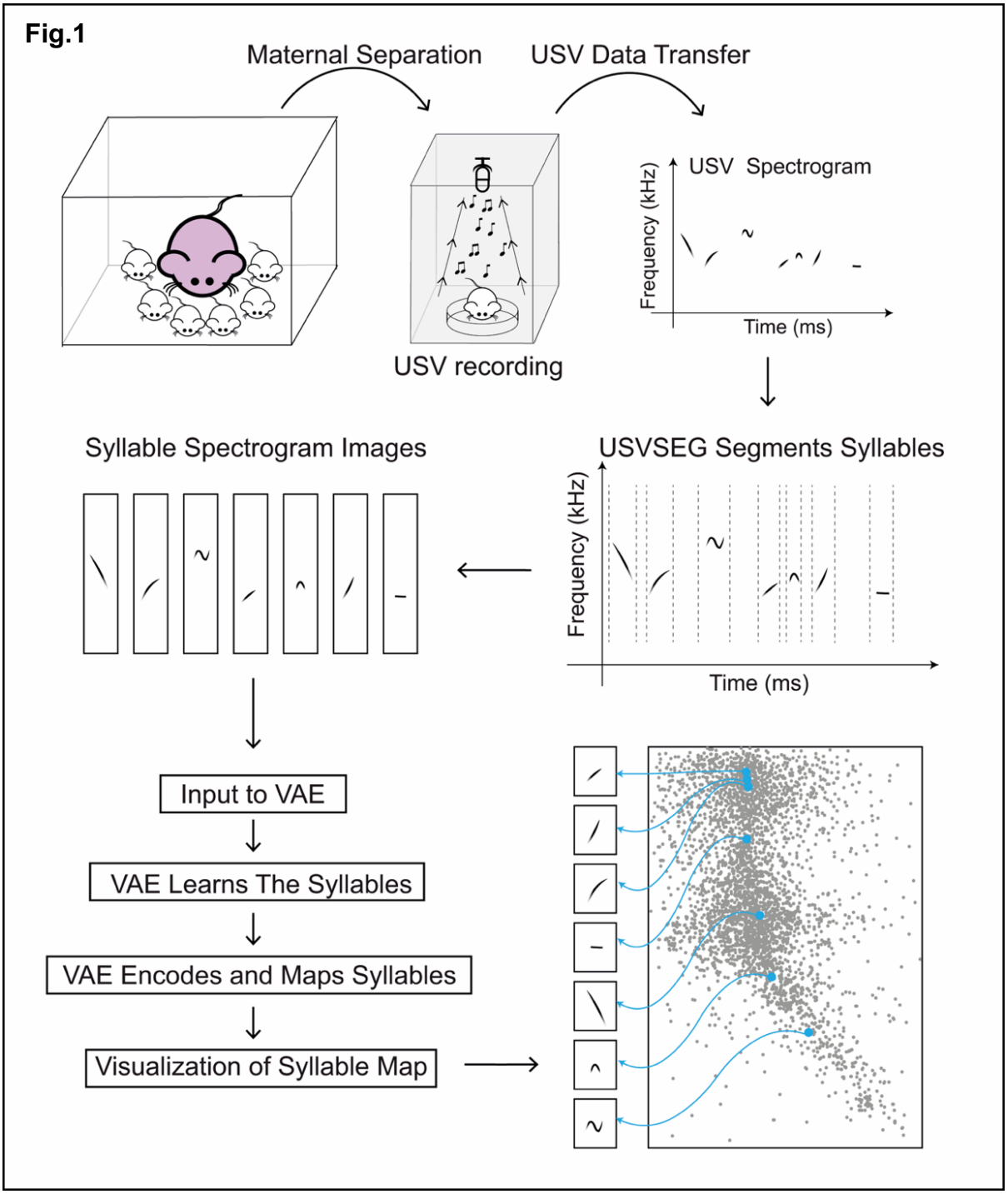


Figure 1. Schematic diagram of the USV analyses by using the VAE. Following maternal separation, USVs were recorded from individual pups one by one. The resulting USVs were processed to draw spectrogram and processed by USVSEG, which segmented the syllables and outputted the syllable images with acoustic parameters. Subsequently, the VAE was trained to learn and encode the syllable images. The resulting syllables were visualized and mapped in a latent space.

153 **Fig.2**

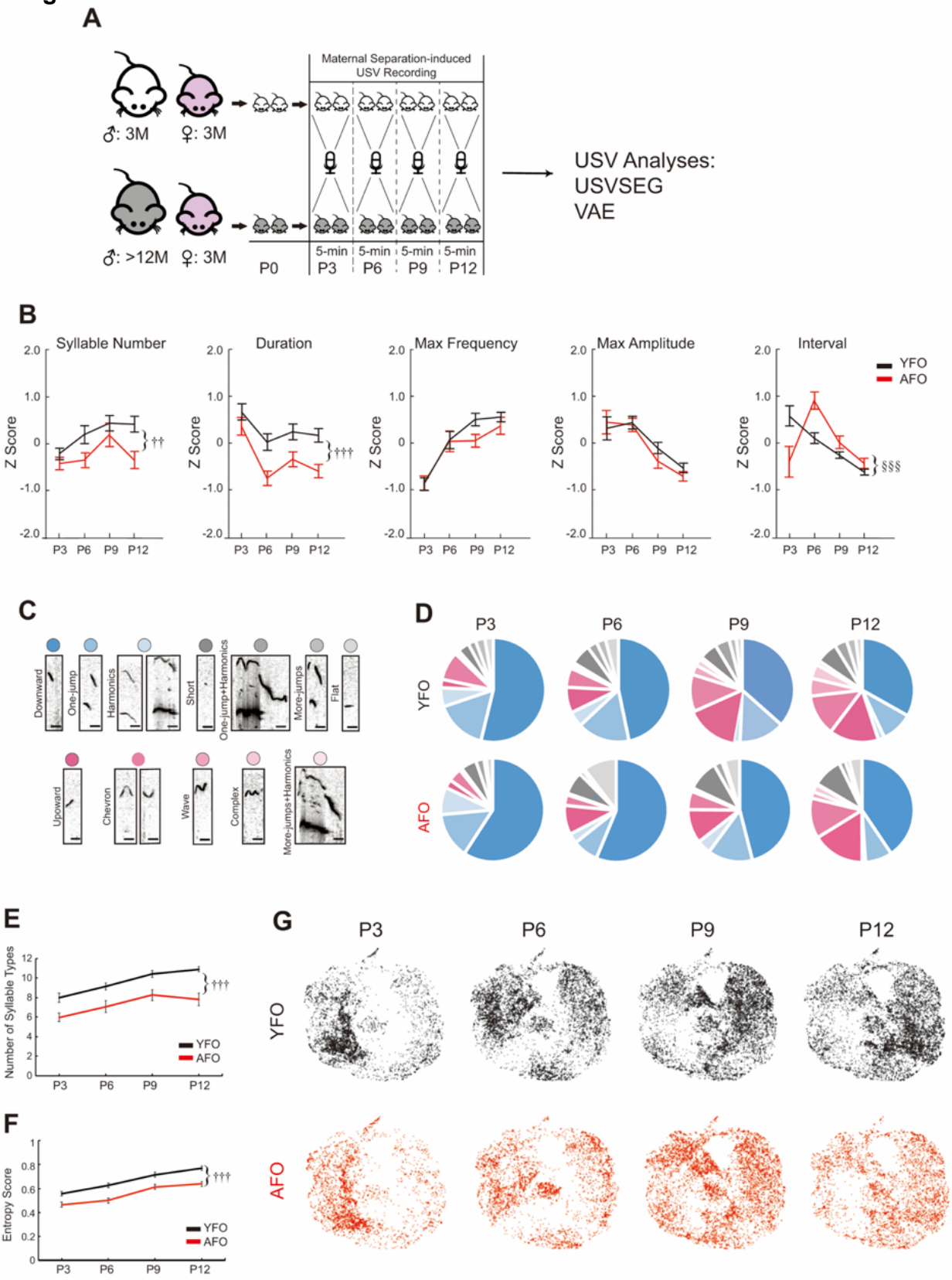

A
Maternal Separation-induced USV Recording
♂: 3M
♀: 3M
♂: >12M
♀: 3M
P0
5-min P3
5-min P6
5-min P9
5-min P12
USV Analyses:
USVSEG
VAE
B
Syllable Number
Duration
Max Frequency
Max Amplitude
Interval
Z Score
YFO
AFO
P3 P6 P9 P12
C
Downward
One-jump
Harmonics
Short
One-jump+Harmonics
More-jumps
Flat
Upward
Chevron
Wave
Complex
More-jumps+Harmonics
D
P3
P6
P9
P12
YFO
AFO
E
Number of Syllable Types
YFO
AFO
F
Entropy Score
YFO
AFO
G
P3
P6
P9
P12
YFO
AFO

Figure 2. Paternal aging alters the acoustic features of syllables from offspring. (**A**) Experimental design. (**B**) 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). (**C**) Typical sonograms of twelve types of syllables obtained from USVSEG. Scale bar = 60 kHz, 20 ms. (**D**) Pie graphs show the percentage composition of the twelve types of syllables in YFO and AFO from P3 to P12. The colors indicate the syllable type shown in (C). Magenta, blue, and grey syllable types indicate those that showed percentage increases, decreases, and no specific change in YFO, respectively. (**E**) Developmental trajectories of syllable types. ††† $p < 0.001$ indicates a significant main effect of the father's age (ANOVA). (**F**) Normalized entropy scores for syllables. ††† $p < 0.001$ indicates a significant main effect of the father's age (ANOVA). (**G**) VAE takes spectrograms from the syllables of YFO and AFO as input, maps, 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.

171 **Table 1. The number of syllables detected from the models of paternal aging**
172 **and *Auts2* mutant.**

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

Fig.3

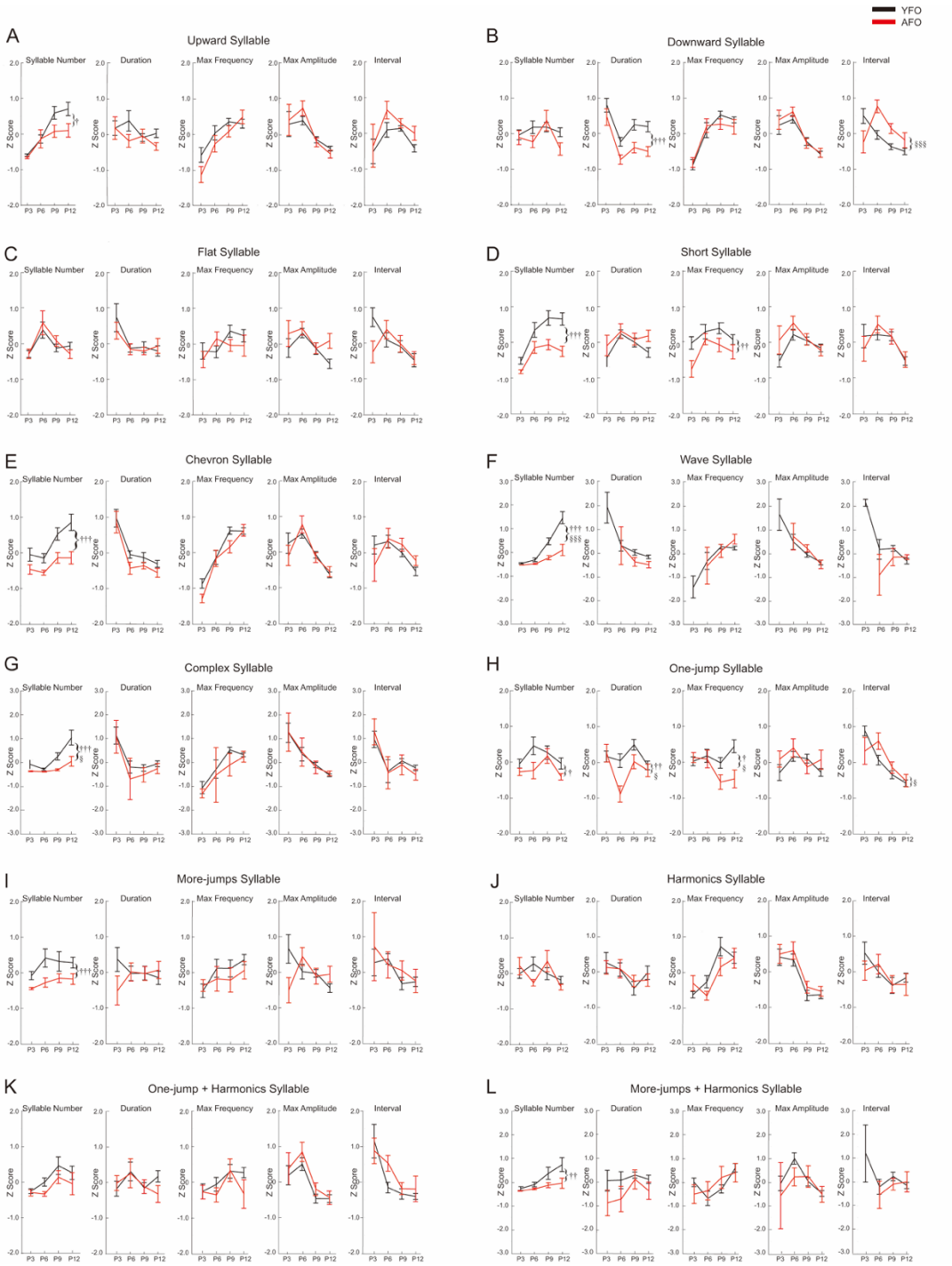


Figure 3. Paternal aging alters the developmental trajectory of distinct types of syllables. The syllable number, duration, maximum frequency, maximum amplitude, and interval of distinct types of syllables from YFO and AFO. (**A**)-(**L**) show the developmental trajectories of five-syllable features among twelve types of syllables. † $p < 0.05$, †† $p < 0.01$, ††† $p < 0.001$ 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).

181 **Supplementary Table 1. Paternal aging affects the percentage composition**
182 **of distinct syllables.**

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 + harmonics	0.86 ± 0.20%: YFO vs 0.20 ± 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

Next, we focused on the diversity of syllables. Our results showed that YFO emitted USVs that contained increasingly more types of syllables with age, while AFO produced USVs with significantly fewer syllable types from P3 to P12 (**Fig. 2E**). To evaluate the difference in diversity of syllable types between the two groups, we calculated entropy scores as an indicator of production uniformity across the syllable types in individual offspring (**Fig. 2F**). As expected, YFO emitted syllables with increased entropy scores during the postnatal period, which is consistent with a previous study (Grimsley et al., 2011). However, AFO showed lower entropy scores than YFO across all postnatal stages. These data indicate that AFO exhibited fewer types and a narrower diversity of syllables, and thus had a poorer vocalization repertoire.

To objectively identify significant USV variations, we next took an unsupervised modeling approach; VAE was used to characterize and quantify subtle changes in behavioral variability. A series of points were mapped in the inferred latent space and visualized (**Fig. 2G**). The number of syllables detected by VAE again demonstrated that 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 syllables at each developmental stage in both YFO and AFO confirmed that USVs developed through a dynamic process in early postnatal stages. Interestingly, while syllable distribution patterns were similar between YFO and AFO at P3, differences became apparent from P6 to P12 and were particularly evident at P12. The major difference, as shown in **Fig. 2G,** was the presence of a distinct aggregated distribution pattern in USVs from YFO that was less obvious in AFO. Missing parts of aggregated syllables may be explained by the reduced number and lack of important syllables that potentially reflect developmental milestones in some atypical individuals. This phenotype indicates a deficiency in vocal development in AFO. The two approaches reached the same conclusion; that is, paternal aging altered syllable development and led to deficiencies in the composition and diversity of USVs.

Paternal aging increases the number of offspring with atypical patterns in syllable development

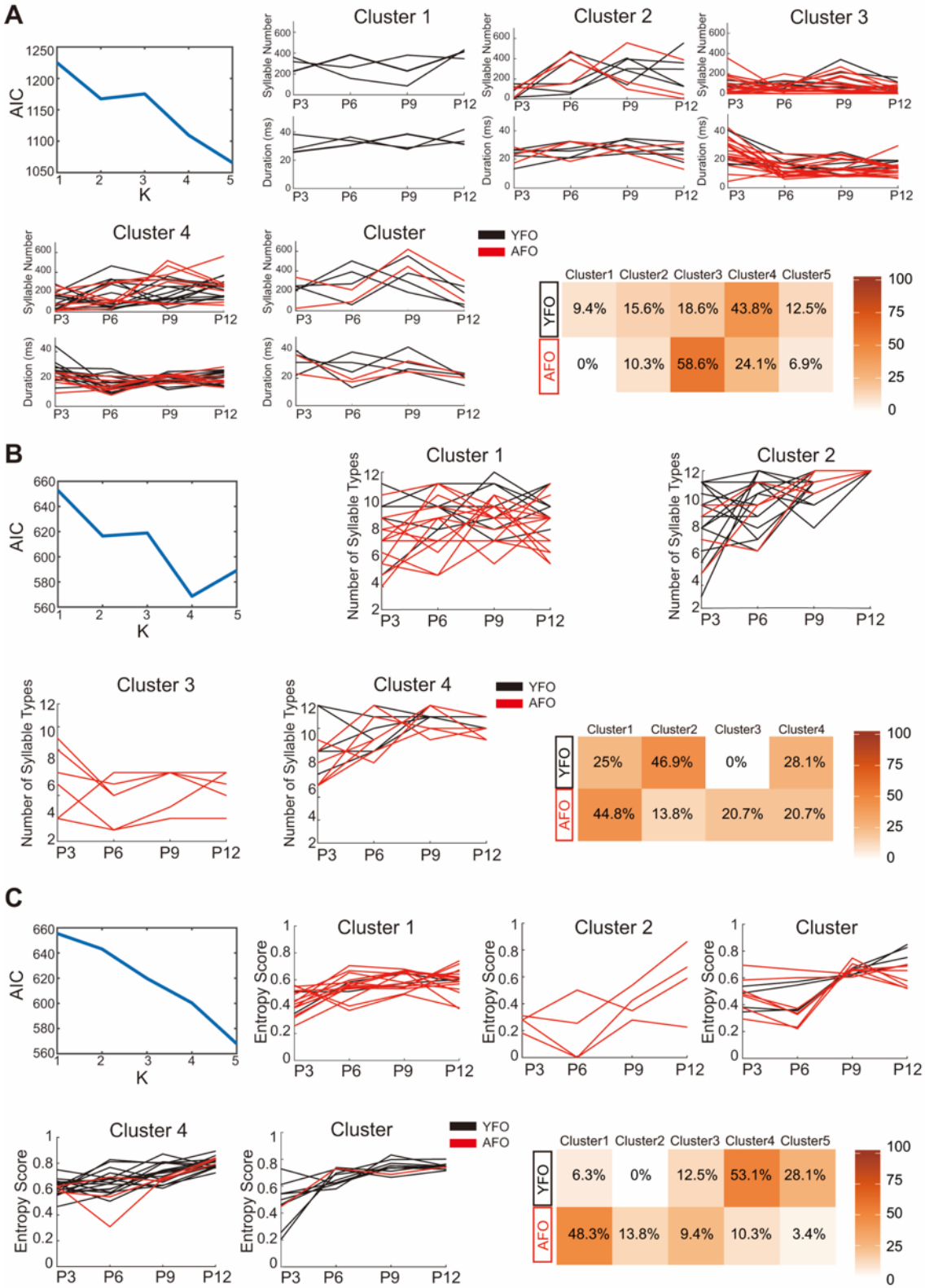
Further, we examined individual differences among pups in the YFO and AFO groups because ASD children often exhibit atypical development. To understand longitudinal syllable development in individuals, based on the USVSEG data, we first performed clustering analyses using Gaussian mixture models. The 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 AFO were significantly decreased in number and duration, we first separated the offspring into clusters based on syllable number and duration (**Fig. 4A**). AIC suggested that five clusters were optimal. The developmental patterns in syllable number and duration observed in YFO were distributed across the five clusters and concentrated in the fourth cluster. In contrast, those in 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 YFO and AFO (p = 0.02). These results show that the dominant developmental patterns of the syllable number and duration were different between YFO and AFO.

We next classified the syllable types into four clusters based on AIC (**Fig. 4B**). Syllables emitted by YFO and AFO were classified into three and four clusters, respectively. Most of the syllables emitted by YFO belonged to the second cluster, while most of those from AFO belonged to the first cluster. Interestingly, the third cluster contained only syllable types emitted by individual AFO. Cluster distribution was significantly different between the two groups (p = 0.003, chi-squared test). Furthermore, to examine the development of syllable diversity in individuals, we clustered normalized entropy (**Fig. 4C**). Five clusters were used based on AIC. Cluster distribution was significantly different between YFO and AFO (p < 0.001, chi-squared test). YFO occupied four clusters and dominated the fourth cluster, while 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 AFO differs from that in YFO. Therefore, paternal aging significantly affected the longitudinal developmental patterns of syllable type and diversity in offspring, with some patterns observed to be unique only to individual AFO.

Fig.4


A
AIC
K
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster
Syllable Number
Duration (ms)
P3
P6
P9
P12
YFO
AFO
Cluster1 Cluster2 Cluster3 Cluster4 Cluster5
YFO 9.4% 15.6% 18.6% 43.8% 12.5%
AFO 0% 10.3% 58.6% 24.1% 6.9%
B
Number of Syllable Types
Cluster1 Cluster2 Cluster3 Cluster4
YFO 25% 46.9% 0% 28.1%
AFO 44.8% 13.8% 20.7% 20.7%
C
Entropy Score
Cluster1 Cluster2 Cluster3 Cluster4 Cluster5
YFO 6.3% 0% 12.5% 53.1% 28.1%
AFO 48.3% 13.8% 9.4% 10.3% 3.4%

249

Figure 4. **Clustering analyses reveal individual-level differences between YFO and AFO.** (**A**) Separation of 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 according to syllable number and duration. The heat map summarizes the percentage of YFO and AFO in each cluster. (**B**) Separation of individual patterns according to syllable type. 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 according to syllable type. The heat map summarizes the percentage of YFO and AFO in each cluster. (**C**) Separation of individual patterns according to 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, according to normalized entropy. The heat map summarizes the percentage of YFO and AFO in each cluster.

To trace the developmental patterns among individuals, we applied PCA analyses to USVSEG data to group the syllable features exhibited by each individual (**Fig. 5A**). At P3, clusters comprising data from 90% of offspring in each group were large, indicating that there were wide individual differences in vocal communication between pups regardless of paternal age. However, there was a minor but significant difference between 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 YFO showed developmental convergence in vocal communication patterns, the variability among AFO remained wide even at P12. We also applied t-SNE to VAE data to analyze the USV repertoire of individual pups (**Fig. 5B**). Again, the vocal repertoires of 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 our PCA findings. We subsequently summarized data from individual traces and highlighted the initial stage (P3) and terminal stage (P12) (**Fig. 5C**). In both PCA and t-SNE, the inferred spatial positions of YFO and AFO were clearly similar at P3 but became significantly different in later developmental stages. Thus, the two approaches produced the same conclusion; the paternal aging led to the misregulation of vocal development and an increase in pups with atypical vocalizations.

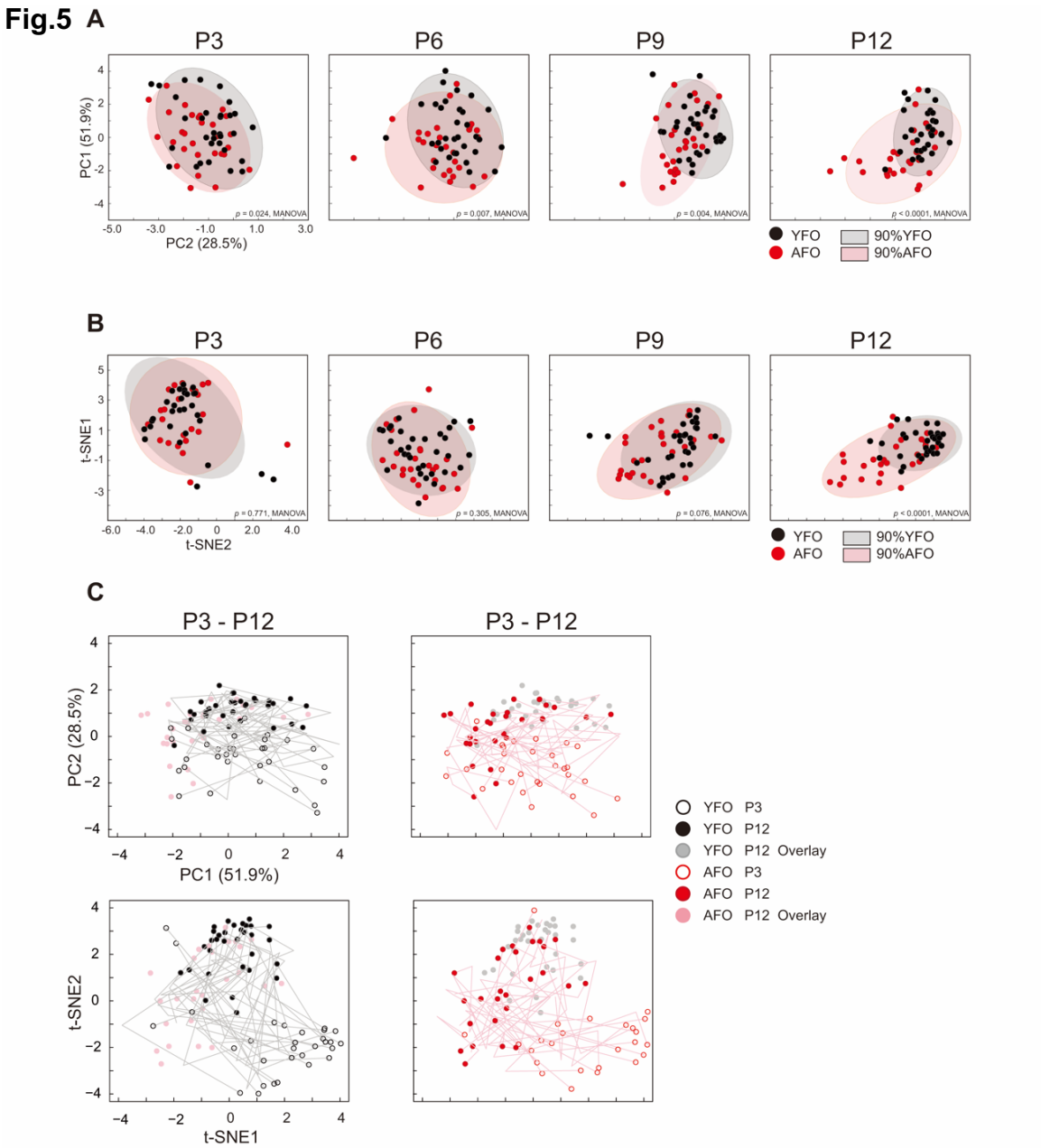


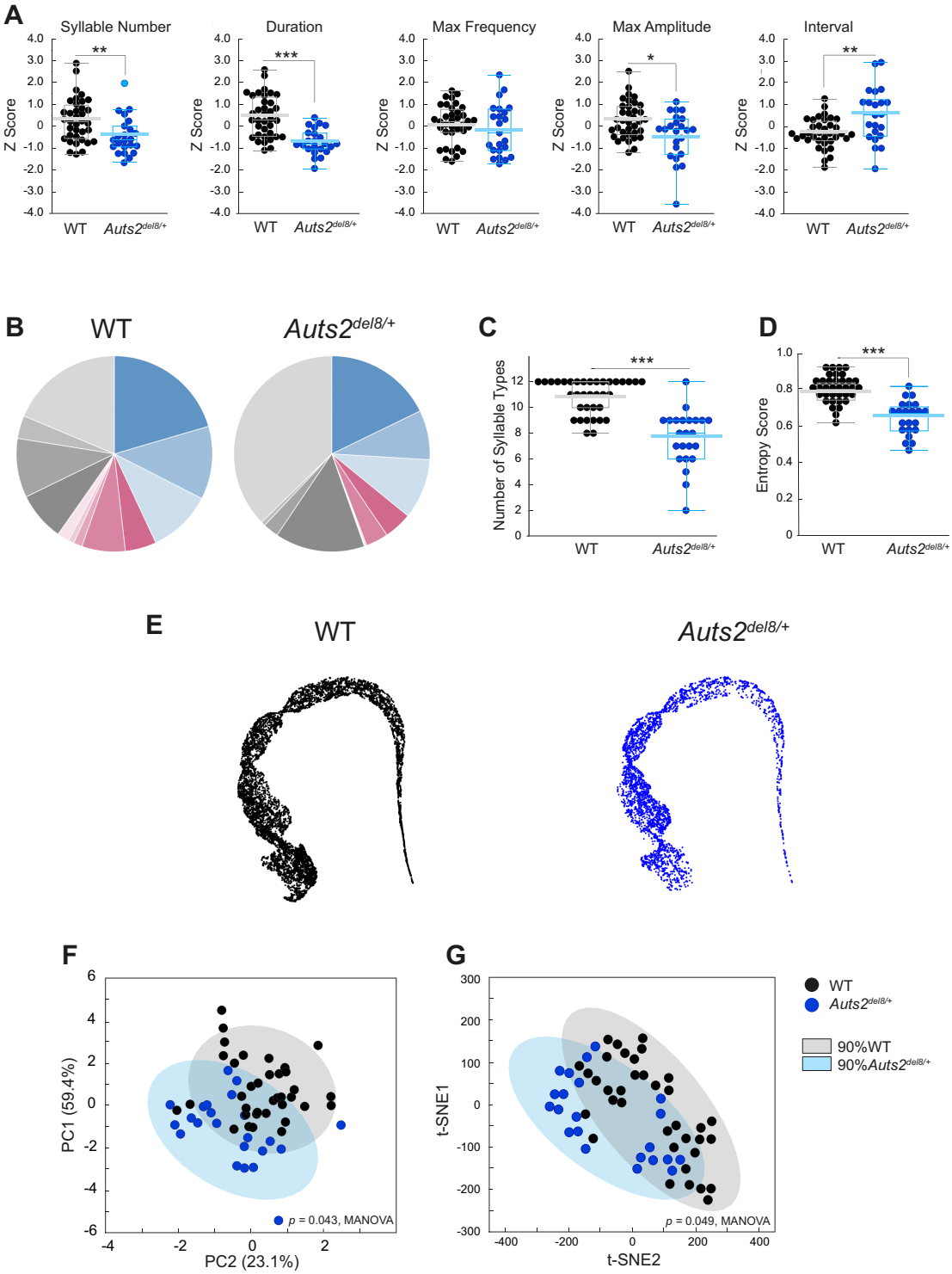
Figure 5. Paternal aging increases the number of offspring with atypical developmental patterns. (**A**) Based on USVSEG, PCA summarized the patterns of individual YFO and AFO from P3 to P12. Each point represents the syllable features of one pup. Shorter distances between points indicate higher similarity in syllable features. Grey and light red clusters contain 90% of the population of YFO and AFO, respectively. (**B**) Based on VAE, t-SNE summarized the patterns of individual YFO and AFO from P3 to 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 YFO and AFO, respectively. (**C**) Summary of individual traces at P3 and P12 based on (**A**) and (**B**). Each point represents the syllable features of one pup.

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

To study syllable phenotypes in a genetic ASD model, the syllables emitted by wildtype and *Auts2*$^{del8/+}$ pups at P7 were recorded (**Table 1)**. Following the procedure of the above, the syllable acoustic features were obtained from USVSEG (**Fig. 6A**). A reduced number, shorter duration, and lower amplitude of syllables, along with longer intervals between adjacent syllables, were revealed in *Auts2*$^{del8/+}$ pups. After syllable classification, the significantly different syllable composition of wildtype and *Auts2*$^{del8/+}$ pups were also detected ($F(11, 47) = 4.483$, $p < 0.001$, MANOVA; **Fig. 6B and Supplementary 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. Correspondingly, the number of syllable types and the entropy scores were also showed significant reductions in the syllables from *Auts2*$^{del8/+}$ pups (**Fig. 6C** and **6D**). Furthermore, VAE mapped the syllables of wildtype and *Auts2*$^{del8/+}$ pups in an inferring space (**Fig. 6E**) that displayed an entirely distinct pattern to those of pups derived from young and aged fathers. As expected, in this inferring syllable map, a significantly different syllable distribution pattern was exhibited between the pups of wildtype and *Auts2*$^{del8/+}$. Some area of syllables, which highly concentrated in the syllable map of wildtype pups, was vanished in the map of *Auts2*$^{del8/+}$ pups.

To evaluate the differences at the individual level, PCA and t-SNE were performed and visualized (**Fig. 6F** and **6G**). 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, in the early postnatal stage, loss of *Auts2* comprehensively impaired the USVs from several aspects such as the quantity, acoustic features, composition, diversity, syllable distribution patterns, and individual patterns.

Fig.6


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/+

329

Figure 6. *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 1C. (**C**) A significant decrease 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 for 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**) VAE takes spectrograms from the syllables of wildtype and *Auts2*$^{del8/+}$ pups as input, maps, 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 USVSEG, PCA summarized the patterns of individual wildtype and *Auts2*$^{del8/+}$ pups. Each point represents the syllable features of one pup. Shorter distances between points indicate higher similarity in syllable features. Grey and light blue clusters contain 90% of the population of wildtype and *Auts2*$^{del8/+}$ pups, respectively. (**G**) Based on VAE, t-SNE summarized the patterns of individual wildtype and *Auts2*$^{del8/+}$ pups. Each point represents the syllable features of one pup. Shorter distances between points indicate higher similarity in syllable features. Grey and light blue clusters contain 90% of the population of wildtype and *Auts2*$^{del8/+}$ pups, respectively.

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

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*$^{del8/+}$, $p < 0.001$

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

Discussion

Previous studies have shown that USVs from mouse pups may be analogous to crying in human infants and are thus used as one of a few tools to understand behavioral development during the early postnatal period (Branchi et al., 2001; Scattoni et al., 2009). Thus, a variety of USV impairments have been observed in distinct genetic models of ASD. For example, a reduction in syllable diversity has been described in *Cd157* KO mice (Lopatina et al., 2017) and *Tbx1*-deficient heterozygous mice (Takahashi et al., 2016); the changed syllable compositions are reported in other genetic ASD mouse models such as the *Reelin* mutant (Romano et al., 2013), *fmr1* knockout (Lai et al., 2014), and *ScSn-Dmd*^{mdx}*/J* mutant (Miranda et al., 2015). Regarding non-genetic factors, advanced paternal age is a risk factor for ASD (Hultman et al., 2011; Kojima et al., 2019; Reichenberg et al., 2006; Sandin et al., 2016), which is modeled in mice by several laboratories including us (Janecka et al., 2015; Sampino et al., 2014; Yoshizaki et al., 2021). However, most of the studies did not pay careful attention to the systematic analyses of vocal communication in these ASD models.

Here, we successfully revealed "atypical" USV development in individual AFO and *Auts2*$^{del8/+}$ pups by comprehensive mathematical analyses. Compared to the behavioral development of children with typical development (so-called "neurotypical"), ASD children show a variety of "atypical" behavioral phenotypes (American Psychiatric Association, 2013; Grzadzinski et al., 2013; Wozniak et al., 2017), exhibiting non-uniformly impaired scores in various criteria. Additionally, these atypical behaviors grow more distinct with age. Some ASD-specific behaviors (e.g., impairment in social cognition, eye contact, language abilities) are not obvious under six months but become gradually evident from the latter part of the first year and second year (Ozonoff et al., 2010; Piven et al., 2017; Zwaigenbaum et al., 2005). Correspondingly, in the model of paternal aging, PCA and tSNE cluster analyses showed unique syllable patterns in individual mouse pups among YFO but more "atypical" patterns among AFO. We also found that while YFO exhibited typical development of vocal repertoires, with differences among individuals gradually converging with age, AFO showed wider variability, becoming most significantly different at the latest postnatal age examined (P12).

Interestingly, our data obtained both from AFO and *Auts2*$^{del8/+}$ pups reveal that the end of the first postanal week, P6-P7, could be a critical window to initiate the 'atypical' patterns in both two models, because the individual differences were revealed by PCA data in both two models at this timing. For an early diagnosis of ASD, our study highlights the significance of indicating subtle "atypical" or "neuro-diverse" behaviors among individuals at early infancy.

The novel approach we presented here for the analyses of USV syllables represents the first attempt to derive unsupervised data-driven developmental trajectories from maternal separation-induced vocalizations in pups. Mapping syllable distribution in an inferring space from a perspective of machine learning clearly showed the distinct features among the two models, which cannot be detected by using the acoustic features obtained from traditional USV analysis. We would like to emphasize also that the semi-automatic method USVSE confirmed and validated the unsupervised VAE in the detection of 'atypical' individuals. It is

yielded the common conclusion that paternal aging and *Aust2* mutant indeed lead to a deficiency in the development of vocal communication in atypical individuals. In addition, the machine learning approach provides us a unique angle to distinguish the potential subtypes of ASD that could be induced by various pathoetiologies. Although we lack comparison analyses with the large literature and other advanced approaches such as nearest neighbors (Kollmorgen et al., 2020), this approach paves the way for future re-analysis of existing databases across species. In order to identify the very first signs, trace the window timing, and stratify the subtypes of ASD, a similar unsupervised learning system of crying analysis can be instrumental in providing a non-invasive and follow-up tool for infants at high risk. The wise use of the machine learning toolkit has thus an incredible potential for new discoveries of translatable key biomarkers.

Methods

Experimental animals

1. To obtain the pups derived from three-month-old (young) or >12-month-old (aged), 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 were selected randomly from 5 young fathers and 29 offspring from 5 aged fathers. The average litter size and number of pups at P0 were 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.
2. The *Auts2*$^{del8/+}$ pups with C57BL/6N background, which exon 8 of the *Auts2* gene has been removed, were generated by *Auts2*$^{del8/+}$ male mice with wildtype female mice as the previous research (Hori et al., 2020). After mating, each female mouse was separated from the male mouse. Totally 8 litters, including 36 wildtype pups and 23 *Auts2*$^{del8/+}$ pups, were obtained. This model of *Auts2* mutant was created at the National Institute of Neuroscience.

Experimental design

At the scheduled postnatal days, pups were separated from their mothers and littermates to record the maternal separation induced USVs (described below). In the mouse model of paternal aging, at postnatal day 3 (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 analyses (mortality rate, YFO: 5.7% vs. AFO: 12.1%). Finally, the detailed USVs analyses were performed and interpreted. This study was not blinded to investigators. 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, respectively. 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, while the pups of *Auts2*$^{del8/+}$ and wildtype were recorded 3 minutes after the first detected call 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. Both male and female pups were analyzed. Room temperature was maintained at 22˚C.

Syllable segmentation and classification

Acoustic waveforms were processed using a GUI-based MATLAB script ("USVSEG") originally developed for segmenting USVs emitted by rodents (Tachibana et al., 2020). Briefly, the script 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. The program script 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 of enhanced spectrograms generated by the MATLAB program script. Ten of these syllable types (#1-10 below) were similar to those described previously (Kikusui et al., 2011; Scattoni et al., 2008). The noise sounds that were mistakenly segmented by the program (e.g., scratching noises) were manually removed from further analyses. Definitions of each of the syllable categories are as follows (also see **Fig. 2C**):

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 variational autoencoder (VAE), an unsupervised machine learning algorithm to extract essential features without any instructional input. We used 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) as the standard-setting. The latent dimension was fixed to 32, which was found by the previous study to be sufficient for all training runs. Because we found that an audio segmentation method implemented in the original program detected large amounts of noise and was unable to detect some subtle syllables, we decided to use USVSEG for audio segmentation before using VAE. USV syllables were identified and segmented by USVSEG using the following parameters: frequency range = 60–120 kHz in the model of paternal aging and 40 – 120 kHz in the model of *Auts2* mutant; minimum gap = 10 ms; minimum duration = 2 ms. After training, 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 requires a minimum of 10 syllables. After removing noise, the VAE was trained again to map single-syllable spectrogram images, such as those shown in **Fig. 2G**. To visualize variations among individuals, t-distributed Stochastic Neighbor Embedding (t-SNE) was applied for each pup (**Fig. 5B-5C** and **Fig. 6G**). 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. Two main effects (i.e., father's age and postnatal day) and their interaction (i.e., father's age × postnatal day) were examined by ANOVA. 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 and 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 as a measure of 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 produced in 5 minutes 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. To understand development at the individual level, clustering analysis with Gaussian mixture models (GMMs) was applied, 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) (Konishi & Kitagawa, 2008; McLachlan & Peel, 2000). Based on the fitted GMM, each individual 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 typical 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. 5A–B, Fig. 6A, 6C-D,** and **6F-G** were constructed using JMP13 Pro software, and **Fig. 5C** was made using R software.

Competing interests

Authors declare no competing interests.

Author Contributions

Designed research: L. M. and N. O.
Developed paternal aging model, collected the USV of 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.
Wrote the paper: L. M., H. I., R. K., V.T., N. O.
Discussed and interpreted the results: all authors

Acknowledgments

General: The authors thank Ms. Sayaka Makino for animal care. The authors also appreciate all members of their laboratories for contributive discussions.
Funding: This work was supported by KAKENHI in Innovative Areas (Grant Number 16H06530) from the Ministry of Education, Culture, Sports, Science, and Technology.

Data availability

All relevant data are available from the authors upon request.

References

Acebo, C., & Thoman, E. B. (1992). Crying as social behavior. *Infant Mental Health Journal*, *13*(1), 67-82.

Acebo, C., & Thoman, E. B. (1995). Role of infant crying in the early mother-infant dialogue. *Physiology and Behavior*, *57*(3), 541-547.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.).* American Psychiatric Association. https://doi.org/10.1176/appi.books. 9780890425596

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, *23*(2), 185-188. https://doi.org/10.1038/13810

Bakkaloglu, B., O'Roak, B. J., Louvi, A., Gupta, A. R., Abelson, J. F., Morgan, T. M., Chawarska, K., Klin, A., Ercan-Sencicek, A. G., Stillman, A. A., Tanriover, G., Abrahams, B. S., Duvall, J. A., Robbins, E. M., Geschwind, D. H., Biederer, T., Gunel, M., Lifton, R. P., & State, M. W. (2008). Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet*, *82*(1), 165-173. https://doi.org/10.1016/j.ajhg.2007.09.017

Berkel, S., Marshall, C. R., 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, S. W., & Rappold, G. A. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet*, *42*(6), 489-491. https://doi.org/10.1038/ng.589

Branchi, I., Santucci, D., & Alleva, E. (2001). Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behav Brain Res*, *125*(1-2), 49-56. https://doi.org/10.1016/S0166-4328(01)00277-7

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. https://doi.org/10.1007/s10519-004-0860-9

Dougherty, J. D., Maloney, S. E., Wozniak, D. F., Rieger, M. A., Sonnenblick, L., Coppola, G., Mahieu, N. G., Zhang, J., Cai, J., Patti, G. J., Abrahams, B. S., Geschwind, D. H., & Heintz, N. (2013). The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *J Neurosci*, *33*(7), 2732-2753. https://doi.org/10.1523/JNEUROSCI.4762-12.2013

Dudek, J., Faress, A., Bornstein, M. H., & Haley, D. W. (2016). Infant cries rattle adult cognition. *PLoS One*, *11*(5), e0154283. https://doi.org/10.1371/journal.pone.0154283

Esposito, G., Hiroi, N., & Scattoni, M. L. (2017). Cry, baby, cry: expression of distress as a biomarker and modulator in autism spectrum disorder. *International Journal of Neuropsychopharmacology*, *20*(6), 498-503. https://doi.org/10.1093/ijnp/pyx014

Goffinet, J., Mooney, R., & Pearson, J. (2019). Inferring low-dimensional latent descriptions of animal vocalizations. *bioRxiv*. https://doi.org/10.1101/811661

Grimsley, J. M., Monaghan, J. J., & Wenstrup, J. J. (2011). Development of social vocalizations in mice. *PLoS One*, *6*(3), e17460. https://doi.org/10.1371/journal.pone.0017460

Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes [Review]. *Mol Autism*, *4*(12). https://doi.org/10.1186/2040-2392-4-12

Hahn, M. E., & Lavooy, M. J. (2005). A Review of the methods of studies on infant ultrasound production and maternal retrieval in small rodents. *Behav Genet.*, *35 (1)*, 31-52. https://doi.org/10.1007/s10519-004-0854-7

Homberg, J. R., Kyzar, E. J., Nguyen, M., Norton, W. H., Pittman, J., Poudel, M. K., Gaikwad, S., Nakamura, S., Koshiba, M., Yamanouchi, H., Scattoni, M. L., Ullman, J. F., Diamond, D. M., Kaluyeva, A. A., Parker, M. O., Klimenko, V. M., Apryatin, S. A., Brown, R. E., Song, C., Gainetdinov, R. R., Gottesman, I. I., & Kalueff, A. V. (2016). Understanding autism and other neurodevelopmental disorders through experimental translational neurobehavioral models. *Neurosci Biobehav Rev*, *65*, 292-312. https://doi.org/10.1016/j.neubiorev.2016.03.013

Hori, K., Yamashiro, K., Nagai, T., Shan, W., Egusa, S. F., Shimaoka, K., Kuniishi, H., Sekiguchi, M., Go, Y., Tatsumoto, S., Yamada, M., Shiraishi, R., Kanno, K., Miyashita, S., Sakamoto, A., Abe, M., Sakimura, K., Sone, M., Sohya, K., Kunugi, H., Wada, K., Yamada, K., & Hoshino, M. (2020). AUTS2 Regulation of Synapses for Proper Synaptic Inputs and Social Communication. *iScience*, *23*(6), 101183. https://doi.org/10.1016/j.isci.2020.101183

Hultman, C. M., Sandin, S., Levine, S. Z., 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. *Molecular psychiatry*, *16*(12), 1203-1212. https://doi.org/10.1038/mp.2010.121

Janecka, M., Manduca, A., Servadio, M., Trezza, V., Smith, R., Mill, J., Schalkwyk, L. C., Reichenberg, A., & Fernandes, C. (2015). Effects of advanced paternal age on trajectories of social behavior in offspring. *Genes Brain Behav*, *14*(6), 443-453. https://doi.org/10.1111/gbb.12227

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50-54.

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*(6), 2524-2532. https://doi.org/10.1093/cercor/bhy122

Kollmorgen, S., Hahnloser, R. H. R., & Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature*, *577*(7791), 526-530. https://doi.org/10.1038/s41586-019-1892-x

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

Lai, J. K., Sobala-Drozdowski, M., Zhou, L., Doering, L. C., Faure, P. A., & Foster, J. A. (2014). Temporal and spectral differences in the ultrasonic vocalizations of fragile X knock out mice during postnatal development. *Behav Brain Res*, *259*, 119-130. https://doi.org/10.1016/j.bbr.2013.10.049

Lopatina, O. L., Furuhara, K., Ishihara, K., Salmina, A. B., & 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*. https://doi.org/10.3389/fnins.2017.00266

Lotspeich, L. J., & Ciaranello, R. D. (1993). The neurobiology and genetics of infantile autism. *Int Rev Neurobiol*, *35*, 87-129.

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models.* (1 st ed.). Wiley-Interscience Press.

Miranda, R., Nagapin, F., Bozon, B., Laroche, S., Aubin, T., & Vaillend, C. (2015). Altered social behavior and ultrasonic communication in the dystrophin-deficient mdx mouse model of Duchenne muscular dystrophy. *Mol Autism*, *6*, 60. https://doi.org/10.1186/s13229-015-0053-9

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

Nardone, S., Sams, D. S., Zito, A., Reuveni, E., & Elliott, E. (2017). Dysregulation of cortical neuron DNA methylation profile in autism spectrum disorder. *Cereb Cortex*, *27*(12), 5739-5754. https://doi.org/10.1093/cercor/bhx250

Ozonoff, S., Losif, A. M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Stainfeld, M. B., & Young, G. S. (2010). A Prospective study of the emergence of early behavioral signs of autism. *J Am Acad Child Adolesc Psychiary*, *49*(3), 256-266.

Piven, J., Elison, J. T., & Zylka, M. J. (2017). Toward a conceptual framework for early brain and behavior development in autism. *Mol Psychiatry*, *22*(10), 1385-1394. https://doi.org/10.1038/mp.2017.131

Reichenberg, A., Gross, R., Weiser, M., Bresnahan, M., Silverman, J., Harlap, S., Rabinowitz, J., Shulman, C., Malaspina, D., Lubin, G., Knobler, H. Y., Davidson, M., & Susser, E. (2006). Advancing paternal age and autism [original article]. *Arch Gen Psychiatry*, *63(9)*, 1026-1032. https://doi.org/10.1001/archpsyc.63.9.1026

Romano, E., Michetti, C., Caruso, A., Laviola, G., & Scattoni, M. L. (2013). Characterization of neonatal vocal and motor repertoire of reelin mutant mice. *PLoS One*, *8*(5), e64407. https://doi.org/10.1371/journal.pone.0064407

Sampino, S., Juszczak, G. R., Zacchini, F., Swiergiel, A. H., Modlinski, J. A., Loi, P., & Ptak, G. E. (2014). Grand-paternal age and the development of

autism-like symptoms in mice progeny. *Transl Psychiatry*, *4*, e386. https://doi.org/10.1038/tp.2014.27

Sandin, S., Schendel, D., Magnusson, P., Hultman, C., Surén, P., Susser, E., Grønborg, T., Gissler, M., Gunnes, N., Gross, R., Henning, M., Bresnahan, M., Sourander, A., Hornig, M., Carter, K., Francis, R., Parner, E., Leonard, H., Rosanoff, M., Stoltenberg, C., & Reichenberg, A. (2016). Autism risk associated with parental age and with increasing difference in age between the parents. *Mol Psychiatry*, *21*(5), 693-700. https://doi.org/10.1038/mp.2015.70

Scattoni, M. L., Crawley, J., & Ricceri, L. (2009). Ultrasonic vocalizations: a tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. *Neurosci Biobehav Rev*, *33*(4), 508-515. https://doi.org/10.1016/j.neubiorev.2008.08.003

Scattoni, M. L., Gandhy, S. U., Ricceri, L., & Crawley, J. N. (2008). Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS One*, *3*(8), e3067. https://doi.org/10.1371/journal.pone.0003067

Schaefer, G. B., & Mendelsohn, N. J. (2008). Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. *Genet Med*, *10*(1), 4-12. https://doi.org/10.1097/GIM.0b013e31815efdd7

Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., & Lester, B. M. (2012). Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Res*, *5*(5), 331-339. https://doi.org/10.1002/aur.1244

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., Schmeidler, J., De Gasperi, R., Sosa, M. A., Rabidou, D., Santucci, A. C., Perl, D., Morrisey, E., & Buxbaum, J. D. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proc Natl Acad Sci U S A*, *102*(27), 9643-9648. https://doi.org/10.1073/pnas.0503739102

Soltis, J. (2004). The signal functions of early infant crying. *Behav Brain Sci.*, *27*, 443-490. https://doi.org/10.1017/s0140525x0400010x

Sultana, R., Yu, C. E., Yu, J., Munson, J., Chen, D., Hua, W., Estes, A., Cortes, F., de la Barra, F., Yu, D., Haider, S. T., Trask, B. J., Green, E. D., Raskind, W. H., Disteche, C. M., Wijsman, E., Dawson, G., Storm, D. R., Schellenberg, G. D., & Villacres, E. C. (2002). Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics*, *80*(2), 129-134. https://doi.org/10.1006/geno.2002.6810

Tachibana, R. O., Kanno, K., Okabe, S., Kobayasi, K. I., & Okanoya, K. (2020). USVSEG: A robust method for segmentation of ultrasonic vocalizations in rodents. *PLoS One*, *15*(2), e0228907. https://doi.org/10.1371/journal.pone.0228907

Takahashi, T., Okabe, S., Broin, P. O., Nishi, A., Ye, K., Beckert, M. V., Izumi, T., Machida, A., Kang, G., Abe, S., Pena, J. L., Golden, A., Kikusui, T., & Hiroi, N. (2016). Structure and function of neonatal social communication in a genetic mouse model of autism. *Mol Psychiatry*, *21*(9), 1208-1214. https://doi.org/10.1038/mp.2015.190

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*. https://doi.org/10.3389/fpsyt.2014.00053

Unwin, L. M., Bruz, I., Maybery, M. T., Reynolds, V., Ciccone, N., Dissanayake, C., Hickey, M., & Whitehouse, A. J. O. (2017). Acoustic properties of cries in 12-Month old infants at high-risk of autism spectrum disorder. *J Autism Dev Disord*, *47*(7), 2108-2119. https://doi.org/10.1007/s10803-017-3119-z

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. https://doi.org/10.1016/j.neubiorev.2014.03.021

Wozniak, R. H., Leezenbaum, N. B., Northrup, J. B., West, K. L., & Iverson, J. M. (2017). The development of autism spectrum disorders: variability and causal complexity. *Wiley Interdiscip Rev Cogn Sci*, *8*(1-2), e1426. https://doi.org/10.1002/wcs.1426

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*(11), e0166665. https://doi.org/10.1371/journal.pone.0166665

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. https://doi.org/10.15252/embr.202051524

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*(11), e0186798. https://doi.org/10.1371/journal.pone.0186798

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J. C., Warren, S. T., & Schlessinger, D. (1991). Fragile X genotype characterized by an unstable region of DNA. *Science*, *252*(5009), 1179-1181. https://doi.org/10.1126/science.252.5009.1179

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *Int J Dev Neurosci*, *23*(2-3), 143-152. https://doi.org/10.1016/j.ijdevneu.2004.05.001

Appendix

Syllable Analyses of USVSEG

1. The USVSEG, which is a MATLAB script, was downloaded from https://sites.google.com/view/rtachi/resources
2. The recorded vocalizations were transferred to an UltraSound Gate 416H detector set (Avisoft Bioacoustics, Germany) at 25–125 kHz and processed to acoustic waveforms.
3. The files of the acoustic waveforms were processed by USVSEG. 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 outputted as an individual segmented spectrogram image with corresponding segmented data (i.e., duration, inter-syllable interval, maximum frequency, and maximum amplitude).
5. The obtained syllable spectrogram images were screened by visual inspection, and the miss-detected syllables' images (i.e., background noise) were excluded manually.
6. After screening, the segmented syllables were manually classified into twelve syllable types and performed further data analyses by using the acoustic parameters.

Syllable Analyses of VAE

1. 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.
2. To reduce noises, we substituted USVSEG's syllable segmenting for AVA's built-in segmenting functions. We used USVSEG for syllable segmentation. The 5-min acoustic waveforms of recorded vocalizations were processed, and syllables were segmented by USVSEG as mentioned above. Following segmenting decisions, the onset/offset time for each syllable was exported as .txt file.
3. The syllable spectrograms were preprocessed through the 'preprocessing step' of the AVA to determine the visualization images when they were fed into the VAE. 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)
```

4. 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.

5. 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
```

6. To manage the large data files and simplify the data analysis, the DataContainer object of AVA was used to bind all the data together. The pups that 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)
```

7. 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')
```