A Data-driven Approach to the Longitudinal Study of Canine Vocal Pattern Development

Hridayesh Lekhak
The University of Texas at Arlington
Arlington, Texas, USA
hxl7195@mavs.uta.edu

Tuan M. Dang
The University of Texas at Arlington
Arlington, Texas, USA
txd0904@mavs.uta.edu

Theron S. Wang
The University of Texas at Arlington
Arlington, Texas, USA
sxw7663@mavs.uta.edu

Kenny Q. Zhu
The University of Texas at Arlington
Arlington, Texas, USA
kenny.zhu@uta.edu

Abstract

Longitudinal studies of animal vocalizations provide crucial insights into developmental patterns and communicative evolution. To aid such investigations in canines, this paper introduces the Canine Age Transition Vocalization Dataset, a large-scale collection of dog vocalizations featuring meticulously verified metadata (including precise birthdate, breed, and individual dog ID) for 125 dogs across 6 common breeds. Our in-depth longitudinal analysis of this dataset then reveals novel findings on how key vocal parameters, encompassing defined bark types and finer-grained acoustic components (Elemental Dog Bark Units, or EDBUs), change as dogs mature. This work, therefore, offers both a significant new resource and foundational data that enable deeper, more nuanced investigations into the lifelong vocal development of dogs and other animal communication.

CCS Concepts

• **Computing methodologies → Machine learning**; **Supervised learning**; **Natural language processing**; **Speech recognition**.

Keywords

Canine Vocalization Dataset; Bioacoustics; Age-Related Vocal Development; Machine Learning; Animal Communication; Longitudinal Study; Bioacoustic Classification

ACM Reference Format:
Hridayesh Lekhak, Tuan M. Dang, Theron S. Wang, and Kenny Q. Zhu. 2025. A Data-driven Approach to the Longitudinal Study of Canine Vocal Pattern Development. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746027.3758175

1 Introduction

Canine cognitive ability changes with age [30, 43]. Addressing the need for longitudinal research in canine vocal communication, this study investigates how bark units and types evolve as dogs mature, offering a nuanced understanding beyond broad age classifications. We hypothesize that dogs, like human infants showing early vocal acoustic development [35], exhibit developmental changes in their vocal repertoire. To explore this, we created a dataset of canine vocalizations from six diverse dog breeds across multiple age groups. Dataset formation involved collecting YouTube and TikTok videos, filtering for desired breeds, extracting initial age information, and processing videos into dog barks.

This dataset supports research on canine vocal evolution using advanced computational methods. Understanding lifelong canine vocal development reveals fundamental aspects of their maturation, social learning, and adaptive communicative strategies. This research is impactful due to dogs' prevalence in human societies. Insights into their age-related developments, including potential vocal links, bear significant implications for canine cognition, health, and behavior—areas with documented age-related changes [31, 45, 62]. We speculate that dogs may develop or modify their vocal units over time, especially in early life. Plausibly, other species show developmental vocal changes: sperm whale calves, for instance, exhibit a 'babbling-like' phase with more diverse codas before acquiring the adult repertoire [23], and wolves show maturation in existing call acoustic features [20]. While African elephants exhibit inter-population structural variations in existing call types rather than new individual vocal unit development [48], cross-species vocal development remains a rich study area. Ultimately, this work advances dog behavior and vocalization research, offering insights into how computational methods can uncover communicative intricacies in dogs and, potentially, other animals.

To ensure clarity in our analysis of canine vocalizations and to avoid potentially misleading analogies with human linguistic structures, we define the following operational terms used throughout this paper.

- **Barkseq:** An audio segment identified by our initial Sound Event Detection (SED) process (detailed in Section 2.4) that contains dog vocalizations. A `Barkseq` may consist of a single vocal event or a sequence of multiple, closely emitted vocal events.
- **Bark Unit (BU):** An individual, continuous vocal sound produced by a dog, representing a distinct vocal utterance. BUs are delineated from surrounding non-vocal segments or adjacent BUs based on acoustic properties such as significant energy dips and low signal variability (detailed in

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758175

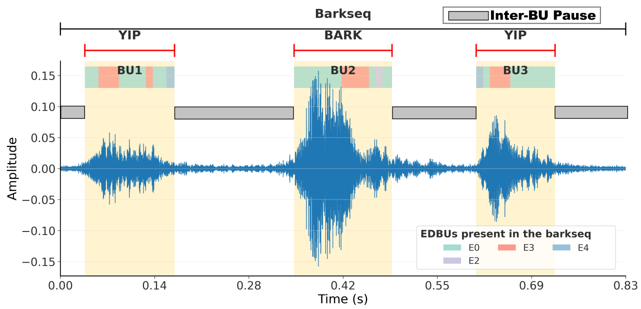


Figure 1: A Barkseq illustrating its BUs, Inter-BU Pauses, and contained EDBUs.

Section 2.5.1) . When multiple BUs occur sequentially within a Barkseq, the interval separating them is termed an Inter-BU Pause (defined below).

- **Inter-BU Pause:** This term refers to the 'silence interval' that creates acoustic separability between Bark Units. It is not necessarily a period of absolute digital silence (i.e., zero signal amplitude), but rather a brief interval between vocal utterances characterized by a distinct reduction in signal energy and substantially lower signal variability (detailed in Section 2.5.2) when compared to the Bark Units themselves.
- **Elemental Dog Bark Unit (EDBU):** An acoustically distinct **sound unit corresponding to a 10ms frame** within an individual **Bark Unit (BU)**. These units are classified into a set of fundamental EDBU *types* (derived per breed by clustering Mel-frequency cepstral coefficient (MFCC) features from such frames, as detailed in Section 3.1.2). Each Bark Unit is thus represented as a sequence of EDBU type labels assigned to its constituent frames.

Each of these defined concepts is visually illustrated in Figure 1.

Our key contributions are:

- We introduce the Canine Age Transition Vocalization Dataset [1], the largest open-source collection of its kind. It has been meticulously curated with precisely verified metadata for 125 dogs from 6 common breeds, including individual dog ID, breed, exact birthdate, and the calculated age group associated with each of the 79,142 bark units (derived from 55,718 Barkseqs totaling 11.4 hours). This rich, diverse dataset, with its detailed longitudinal tracking, is specifically designed to enable robust research into canine vocal development.
- Our primary analytical contribution is a comprehensive longitudinal study utilizing this dataset. We present novel findings on significant developmental changes and age-related trends in canine vocal patterns, including bark unit characteristics, defined bark type usage, and the distribution of Elemental Dog Bark Units (EDBUs) across five distinct life stages and multiple breeds. This work offers the first such comprehensive analysis of dog vocal development, as further discussed and compared with existing literature (see Section 4).

[1] For requesting access to the dataset, please visit: https://github.com/Lekhak123/A-Data-driven-Approach-to-the-Longitudinal-Study-of-Canine-Vocal-Pattern-Development

2 Dataset Creation

On social media platforms such as TikTok and YouTube, there is a significant number of dog owners who regularly post videos of their pet dogs. One active channel may contain hundreds of video clips of a dog spanning a number of years, and some channels or videos include birth date or age information of their dogs. This gives us the opportunity to download videos with dog barks that carry age information. The creation of the entire dataset involves a few steps: *(1) Seed Video Collection for Channel Discovery*, *(2) Valuable Channel Identification*, *(3) Age Determination*, *(4) Audio Preprocessing*, and *(5) Bark Unit Segmentation.*

2.1 Seed Video Collection for Channel Discovery

Our process for identifying suitable YouTube and TikTok channels begins with an initial 'seed video collection' phase. We use targeted, age-related search queries (e.g., "*Shiba Inu turned X years old*") to discover these first-pass videos. The primary purpose of using such queries for these seed videos is not for immediate age annotation, but rather to efficiently identify an initial cohort of videos that, in turn, lead us to the channels that published them. Channels discovered through this method are more likely to contain explicit age mentions or longitudinal content, making them strong candidates for our study. These identified channels then proceed to the 'Valuable Channel Identification' stage (detailed in Section 2.2). Such queries may return many invalid videos. To filter out invalid videos, we trained a binary classifier using a combination of a BERT [16] model and a ViT [17] model. The ViT model performs image classification, taking YouTube video thumbnails as its input, while the BERT model uses a combination of textual metadata such as the title, description, and comments as its input. The accuracy of the BERT model was 95.4%, and the accuracy of the ViT model was 92.6%. We used the agreement of both models to filter out videos. Through this process, we were able to generate a list of high-quality videos about the target dog breeds.

2.2 Valuable Channel Identification

Our initial pool consisted of 7,243 YouTube and TikTok channels, all of which post dog-related videos. From these channels, we selected high-quality candidates for further analysis. Channels were required to demonstrate: (1) **Channel Activity:** uploads spanning more than one year, and (2) **Video Upload Count:** at least 10 videos within that period. These criteria ensured sufficient data for longitudinal tracking and cost-effective age annotation, as identifying a dog's age in one video from an active channel allows annotation for all its videos.

This filtering yielded 950 channels. To identify those containing precise birth date or age information, we employed Llama 3.2 3B [27], a large language model (LLM) from Meta, to scan video titles, descriptions, and channel 'About' sections for relevant keywords or contextual clues. A validation test on 100 channels showed that Llama correctly identified users with age-related information in 40% of cases; discrepancies often arose from mentions of age groups rather than specific ages, leading to model hallucination. For channels Llama flagged as lacking birthdate cues, this assessment was correct nearly 99% of the time.

The LLM shortlisted 451 channels. We then manually verified the dog’s birthdate for these channels by examining ‘About’ sections, birthday celebration videos, or the owner’s social media profiles, including only channels with explicitly provided birthdates. This rigorous process resulted in 125 unique channels, corresponding to 125 individual dogs, as we intentionally avoided users with multiple pets.

The primary strength of this dataset is the pinpoint accuracy of the verified birthdates, crucial for precise age identification. The dataset comprises **125 dogs** from breeds including Shiba Inu (Shiba), Husky, German Shepherd (GSD), Pitbull (Pit), Labrador (Lab), and Chihuahua (Chi).

2.3 Age Determination

To determine a dog’s age in a given video, we first establish an accurate birth date for the dog or its precise age in at least one video. With this anchor, we can infer the dog’s age in other videos from the same source by using the video’s date. For timestamping video content, we prioritize the ‘upload date’ (e.g., as displayed on YouTube) over the metadata ‘creation date’. However, on platforms like YouTube and TikTok, video metadata such as the “creation date” can be unreliable, as it may be altered if the owner modifies the video, rendering the upload date a more stable indicator of when the content became publicly current.

Table 1: Breed-Specific Age Group Timespans.

Breed	P (M)	J (M)	Ado (Y)	Ad (Y)	S (Y)
Chi (Small)	0–5	6–12	1–5	5–11	11+
Shiba (Small)	0–5	6–12	1–5	5–11	11+
Pit (Med)	0–8	9–18	1.5–4	4–9	9+
Husky (Med)	0–8	9–18	1.5–4	4–9	9+
GSD (Large)	0–14	15–24	2–3	3–8	8+
Lab (Large)	0–14	15–24	2–3	3–8	8+

Note: M/Y in headers = Months/Years. `P`, `Ad`, `S` stages align with Embarkvet size-based lifestages (Small, Med, Large). `J`, `Ado` subdivide Embarkvet’s ‘Young Adult’.

Extensive literature highlights significant inter-breed variability in maturation and lifespan, underscoring the need for breed-specific age categorizations in canine research, as universal groupings can obscure developmental phenomena [30, 34, 41, 44, 47, 58]. Such variability is well-documented across diverse breeds [2–4, 6, 10, 12]. Veterinary guidance further emphasizes that size and breed-typical lifespans are crucial for defining lifestages [44, 47].

Consequently, our study on canine vocalizations defines five age categories: `Puppy` (P), `Juvenile` (J), `Adolescent` (Ado), `Adult` (Ad), and `Senior` (S). This model is informed by size-dependent lifestage progressions like those from Embarkvet [47], with our `P`, `Ad` (correlating to Embarkvet’s ‘Mature Adult’), and `S` stages aligning with their Small, Medium, and Large breed guidelines (see Table 1). For finer granularity in analyzing vocal development, our `J` and `Ado` categories subdivide Embarkvet’s broader ‘Young Adult’ phase, facilitating a detailed examination of the transition from puppyhood to full maturity. This tailored `J` and `Ado` delineation aims to reveal nuanced developmental vocal patterns. Table 1 details the precise age ranges for each category and breed, using common breed abbreviations detailed in the table note.

2.4 Audio Preprocessing

We segment dog bark clips from raw video files following the pipeline of Wang et al. [59]. Raw audios are first denoised using AudioSep [40], a language-queried audio source separation tool. To detect dog barks in audio files, we utilize a sound event detection (SED) framework employing BEATs [9] as the audio encoder. After manually labeling 9,000 seconds of dog bark data, our SED model utilizing BEATs achieved an F1 score of 0.8556 on the test set. The model was trained for 5.5 hours on two Nvidia RTX 4090 GPUs with a batch size of 4. This trained model is then used to detect dog Barkseqs from raw videos.

2.5 Bark Unit Segmentation

Our granular analysis of canine vocal development utilizes defined bark types (Section 3.1.1) and Elemental Dog Bark Units (EDBUs) (Section 3.1.2) as primary units. Since a single *Barkseq* from our initial Sound Event Detection phase (Section 2.4) can contain multiple Bark Units (BUs), potentially of different bark types, isolating the core BUs within each *Barkseq* is necessary. We achieve this using a simple segmentation algorithm with a hybrid, dynamic, amplitude-based thresholding approach (detailed in Sections 2.5.1 and 2.5.2). Figure 2 illustrates the result of using this algorithm on two different Barkseqs.

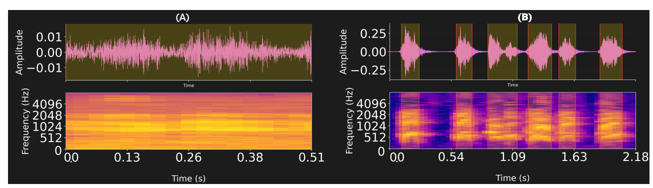


Figure 2: Results of the Bark Unit (BU) segmentation algorithm on two distinct *Barkseqs*. Panel (A) and Panel (B) each display a waveform and its corresponding spectrogram. Detected BU boundaries are indicated by red vertical lines, and the BU regions are highlighted in yellow.

2.5.1 Boundary Identification and Initial Segmentation. Segmentation first identifies inter-BU pauses, characterized by reduced signal variability and energy. To detect BU boundaries, we compute a local signal variability feature, $F_{\text{std}}(k)$ (Eq. 1)—the standard deviation of amplitudes in overlapping frames ($L_{\text{feat}} = 1024, H_{\text{feat}} = 256$) of the input audio $\mathcal{A}_{\text{proc}}(t)$ ($f_s = 16\,\text{kHz}$). This $F_{\text{std}}(k)$ is smoothed with a Gaussian filter ($\sigma_{\text{smooth}} = 2.0$) to $F_{\text{smooth}}(k)$ (Eq. 2), reducing noise and highlighting sustained variability changes, following established audio segmentation techniques [1, 24]. BU boundaries (B_{cand}) are hypothesized at local minima in $F_{\text{smooth}}(k)$, identified via peak-finding on $-F_{\text{smooth}}(k)$ ($p_{\text{prominence}} = 0.0001$). These candidates are validated by confirming an energy drop using a local amplitude envelope $E(t)$. $E(t)$ is computed from frames ($L_{\text{env}} = 256, H_{\text{env}} = 64$) using maximum absolute amplitude (Eq. 3) and smoothed with a median filter ($M_{\text{env}} = 5$). A candidate b at t_b is validated if $E(t_b)$ is significantly lower ($r_{\text{local\_dip}} = 0.3$) than peak amplitudes in its vicinity (Eq. 4), a common heuristic in VAD [18, 52]. If no valid boundaries are found, the entire Barkseq is treated as a single BU. Initial BUs (B_{initial}) are delineated as the audio segments between validated boundaries $\{t_b | b \in b_{\text{valid}}\}$, augmented with Barkseq start (t_{start}) and end (t_{end}) times.

$$F_{\text{std}}(k) = \sqrt{\frac{1}{L_{\text{feat}}} \sum_{i=1}^{L_{\text{feat}}} (x_{k,i} - \bar{x}_k)^2} \tag{1}$$

$$F_{\text{smooth}}(k) = (F_{\text{std}} * G)(k) = \sum_{j=-\infty}^{\infty} F_{\text{std}}(j) \cdot G(k-j; \sigma_{\text{smooth}}) \tag{2}$$

$$E_{\text{raw}}(j) = \max_{m \in Y_j} |\text{sample}_m| \tag{3}$$

$$\begin{aligned} &(E(t_b) < r_{\text{local\_dip}} \cdot \max_{t \in [t_{\text{prev}}, t_b]} E(t)) \\ \wedge \quad &(E(t_b) < r_{\text{local\_dip}} \cdot \max_{t \in [t_b, t_{\text{next}}]} E(t)) \end{aligned} \tag{4}$$

2.5.2 BU Refinement. Initial BUs ($B_i \in B_{\text{initial}}$), whether from detected Inter-BU Pauses or overall `Barkseq` limits, are refined by trimming leading/trailing low-energy segments. This refinement aims to capture the core bark units by removing peripheral, non-core segments (e.g., Inter-BU Pause portions, initial/final `Barkseq` silences) that exhibit Inter-BU Pause characteristics (low energy/variability, as defined in Section 1). For each B_i's audio $\mathcal{A}_{B_i}(t')$, a segment-specific local amplitude envelope $E_{\text{local}}(t')$ is computed (using $L_{\text{env}}, H_{\text{env}}, M_{\text{env}}$ parameters identical to $E(t)$'s). A trimming threshold, $Th_{\text{trim}}(B_i)$, is defined as $r_{\text{trim}} = 0.2$ times the peak of this local envelope, $E_{\text{local, peak}}(B_i)$ (Eqs. 5, 6). The segment is then trimmed to span from the first point t'_{start} to the last t'_{end} where $E_{\text{local}}(t') \geq Th_{\text{trim}}(B_i)$. This use of relative thresholding, adapting to local peak energy, aligns with common speech/audio processing [33, 51] and bioacoustic syllable segmentation techniques [53]. Any segment with duration $< d_{\text{min}} = 0.05$ s (before or after trimming) is discarded.

$$E_{\text{local, peak}}(B_i) = \max_{t' \in B_i} E_{\text{local}}(t') \tag{5}$$

$$Th_{\text{trim}}(B_i) = r_{\text{trim}} \cdot E_{\text{local, peak}}(B_i) \tag{6}$$

Following BU refinement, a final filtering step ensures high-quality data for subsequent analyses. This is performed using a pre-trained Audio Spectrogram Transformer (AST) model (`MIT/ast-finetuned-audioset-10-10-0.4593` [26]). Only BUs with an AST model confidence score of 5% or greater for a 'Dog' event are retained, ensuring clean, high-confidence dog vocalizations (Section 3.1.1 and Section 3.1.2). The complete pipeline, including this filtering, yielded 79,142 individual BUs for the final dataset. The segmentation's effectiveness was assessed by two independent human testers on a randomly sampled set of 100 *Barkseqs*, which included 222 BUs. Testers, examining both waveforms and spectrograms (similar to Figure 2) to identify inter-BU pauses, rated the algorithm's ability to isolate easily distinguishable Bark Units on a 1 (Very Poor) to 5 (Perfect) scale. A perfect score indicated correct isolation of all BUs within a *Barkseq*. A score of 4 denoted successful isolation but with suboptimal trimming of peripheral low-energy segments, while lower scores indicated significant issues like over- or under-segmentation. As detailed in Table 2, the testers' overall scores were 84.6% and 82.3%, with an inter-tester agreement of 97.7%. While a comprehensive comparison to other algorithms was not performed because our approach adapts established audio processing techniques for our dataset's specific characteristics, this human validation confirmed its suitability for the subsequent longitudinal analysis.

Table 2: Results of Human Evaluation by Two Testers for the Bark Unit Segmentation Algorithm.

Metric	Tester 1	Tester 2
Overall Score (%)	84.6	82.3
Inter-Tester Agreement		**97.7%**

Table 3: Final dataset statistics after Bark Unit segmentation and filtering. The table includes the total number of Barkseqs, Bark Units (BUs), the total duration of BUs, and the number of unique dogs and total BUs per age group for each breed.

Metric	Chi	GSD	Husky	Lab	Pit	Shiba
Barkseqs count	5,854	5,557	19,236	8,076	8,355	8,640
BUs count	7,511	9,312	27,287	10,979	12,306	11,747
BUs total dur(h)	0.9	1.1	6.0	1.0	1.1	1.3
Age Group: No. of Unique Dogs (Total BUs)						
(From 125 unique dogs, each can appear in multiple age groups)						
P	6 (328)	8 (1,852)	3 (529)	15 (3,743)	4 (639)	9 (354)
J	10 (477)	7 (1,863)	3 (840)	18 (3,924)	5 (500)	11 (1,105)
Ado	26 (3,537)	8 (1,178)	5 (4,874)	15 (2,608)	6 (1,118)	22 (5,657)
Ad	13 (1,188)	9 (4,262)	7 (16,754)	12 (606)	6 (3,789)	17 (3,846)
S	7 (1,981)	2 (157)	5 (4,290)	1 (98)	3 (6,260)	3 (785)

3 Results and Analysis of Vocalization Patterns

This section analyzes canine vocalizations, focusing on how defined bark types (Subsection 3.2), Elemental Dog Bark Units (EDBUs) (Subsection 3.3), and bark sequence characteristics (Subsection 3.4) change across different life stages and breeds.

3.1 Vocalization Units for Analysis

Our longitudinal analysis tracks changes in two primary vocalization unit levels: defined bark types and finer-grained elemental dog bark units (EDBUs).

3.1.1 Bark Type Definitions. Our bark type definitions are based on the AudioSet ontology [22]. We adopted its categories for dog vocalizations (e.g., Howl, Growl, Whimper, Yip), consolidating AudioSet's 'Bark' and 'Bow-wow' into a single 'Bark/Bow-Wow' type due to their acoustic similarity. To classify the filtered Bark Units (BUs) into these defined types, we utilize the same AST model (as discussed in 2.5.2). Each BU is then assigned the single bark type for which this AST model provides the highest confidence score. The model's classification performance on our dedicated test set for these bark types is summarized in Table 4.

The performance metrics presented in Table 4 were derived from a dedicated test set. This set consisted of 306 Bark Units (BUs), randomly sampled from our broader dataset. Each BU in this test

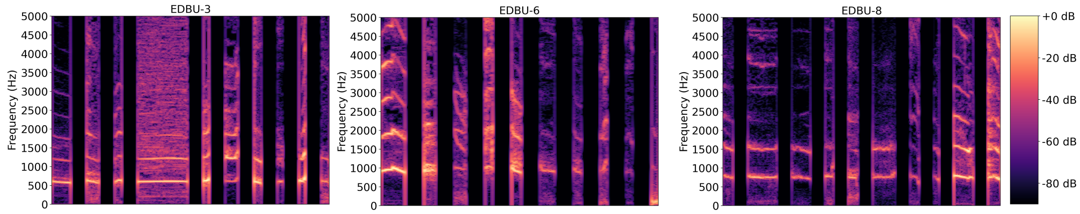


Figure 3: Representative Chihuahua EDBUs (EDBU-3, EDBU-6, EDBU-8): Each spectrogram, formed by concatenating multiple segments of the specified EDBU type from various dogs and Bark Units, showcases consistent acoustic patterns within that EDBU type and highlights spectral differences between distinct EDBU types.

Table 4: Per-Class Classification Performance for Bark Types using AST Model ($N = 306$).

Bark Type	P (%)	R (%)	F1 (%)	Support
Bark/Bow-Wow (B/B-w)	92.2	79.7	85.5	59
Growling (Gr)	96.0	65.8	78.0	73
Howl (Hw)	80.3	100.0	89.1	57
Whimper (Wm)	72.3	82.5	77.0	57
Yip (Yp)	73.9	85.0	79.1	60
Accuracy			***81.7%***	*306*
Macro Avg.	82.9	82.6	81.7	306
Weighted Avg.	83.6	81.7	81.5	306

set was manually assigned a ground truth label following a careful auditory review of both the isolated unit and its parent Barkseq to ensure contextual accuracy. The AST model's predictions for these BUs were then evaluated against these curated ground truth labels to generate the reported performance.

3.1.2 Elemental Dog Bark Units (EDBUs). Beyond categorical bark types, our study investigates finer, breed-specific acoustic components, which we term **Elemental Dog Bark Units (EDBUs)**. To derive EDBUs, we processed up to 4000 randomly selected individual **Bark Units (BUs)** per breed. From short audio frames within these vocalizations (16 kHz sampling rate), we extracted 39 Mel-frequency cepstral coefficient (MFCC) features [14] (window: 25 ms, hop: 10 ms, 128 Mel bands, 512-sample FFT).

Gaussian Mixture Models (GMM) [8] were subsequently applied to these **39-dimensional MFCC features** to identify distinct EDBU types. For each breed, we explored potential cluster numbers (k) from 4 to 20. The optimal k for EDBU clustering (summarized in Table 5) was selected based on the configuration achieving the lowest Davies-Bouldin (DB) Index [13]. The DB Index measures intra-cluster similarity relative to inter-cluster separation, aiding in identifying compact and distinct EDBU clusters.

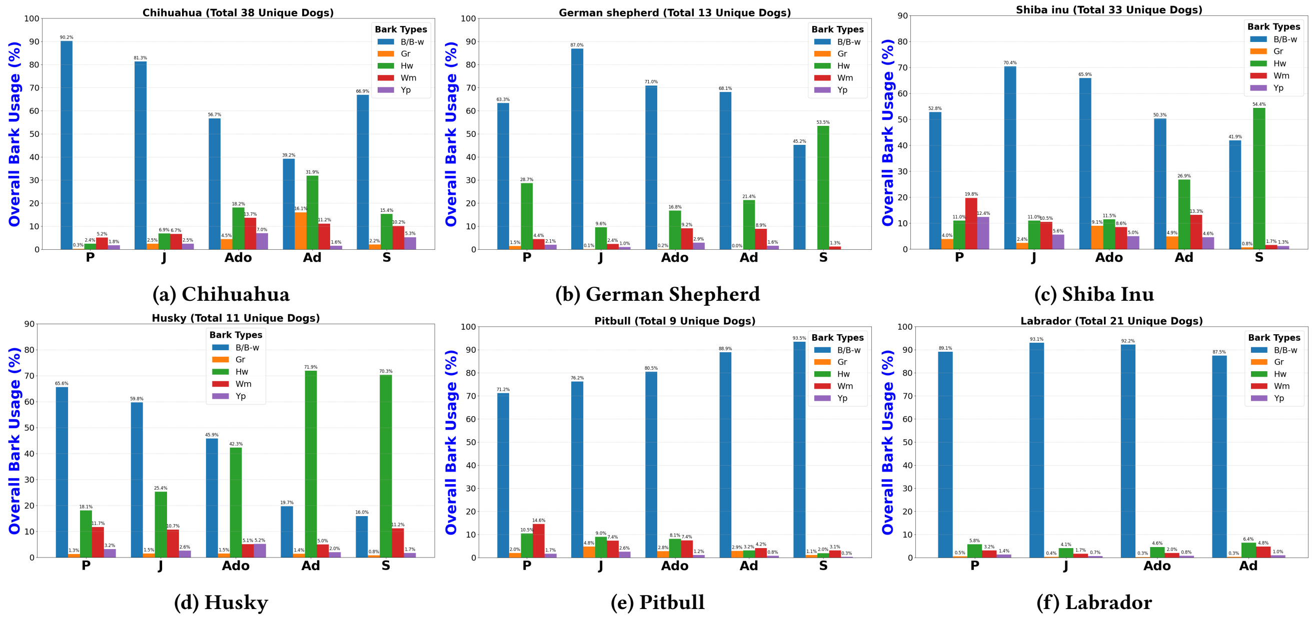
Figure 3 showcases representative acoustic segments for three distinct EDBU types (EDBU-3, EDBU-6, and EDBU-8) found in the Chihuahua breed. While a single EDBU fundamentally corresponds to a 10ms acoustic frame, for clearer visualization in this figure, the segments presented are EDBU sequences. This collage was created by first randomly selecting these three EDBU types. Then, for each selected EDBU type, we identified multiple Bark Units, sourced from different Chihuahua dogs, that contained continuous sequences of at least five 10ms frames classified as that specific EDBU type. To visualize the intra-EDBU type acoustic similarity, these individual EDBU sequences (segments) were concatenated and displayed as a continuous spectrogram within each panel of the figure. A brief pause, represented by a black vertical band in the spectrogram, denotes the boundary between these originally separate, concatenated EDBU segments. The spectrograms display frequencies up to 5000 Hz and are presented side-by-side to facilitate comparison. This approach allows for an examination of both the acoustic consistency within each EDBU type and the clear spectral differences between the distinct EDBU types. Notably, the **formant positions**—visible as bands of concentrated energy—appear **consistent for segments representing the same EDBU type**, despite originating from different Bark Units and dogs. Conversely, these formant structures exhibit **clear differences when comparing one EDBU type to another**, underscoring how the clustering of MFCC features effectively captured the distinct acoustic characteristics of these 10ms EDBU frames.

Table 5: Optimal number of GMM clusters (k) for each dog breed, determined using the Davies-Bouldin Index (lower is better).

Metric	Chi	GSD	Husky	Lab	Pit	Shiba
GMM k	11	9	14	5	6	15
DB Index	3.24	3.61	2.83	3.37	3.08	3.22

3.2 Dog Bark Type Analysis

The overall usage of predefined bark types across five age groups is shown in Figure 4. These bar charts provide a consolidated view of vocalization patterns for each breed, with data sparsity in some age groups addressed through aggregation. Preliminary analysis reveals distinct patterns in bark type usage across breeds and age groups. For Huskies, Figure 4d indicates a visually distinct trend in vocal development, showing a marked increase in the proportion of howling as they mature. This pattern is further detailed in the 2-month binned heatmap presented in Figure 5, where howling appears to become the predominant bark type in older Huskies relative to other vocalizations, particularly from the adolescent stage onwards. The observed increase in howling corresponds with data from the bark sequence duration analysis (Figure 7a), which shows a consistent and sharp rise in average sequence duration for Huskies from puppyhood through to their senior stage. This



(a) Chihuahua (b) German Shepherd (c) Shiba Inu

(d) Husky (e) Pitbull (f) Labrador

Figure 4: Bar plots showing the micro-averaged percentage of bark type usage by age group for each breed.

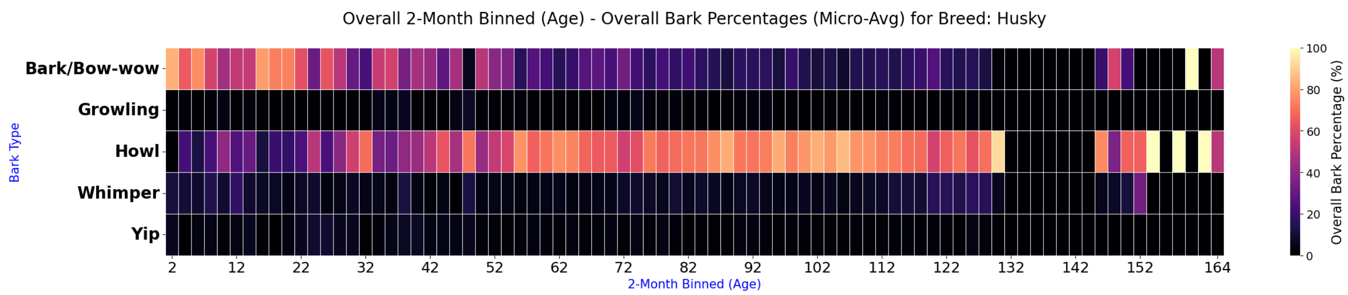


Figure 5: Husky breed: Heatmap of micro-averaged overall bark percentages for various bark types from 11 unique dogs, binned by 2-month age intervals (0–164 months). Black columns denote month-bins with no data.

relationship is consistent with howls being characteristically longer, continuous vocalizations.

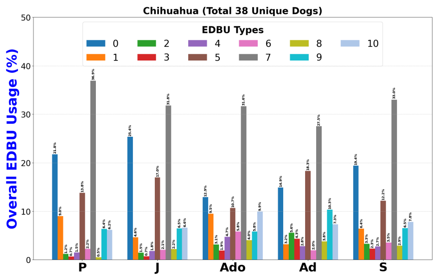
3.3 Elemental Dog Bark Unit (EDBU) Analysis

Preliminary observations from Figure 6 indicate breed-specific EDBU distribution patterns across life stages, with the proportional usage of many EDBU types appearing to expand or shift after puppyhood, potentially suggesting a phase of learned vocal development as their repertoires diversify. For instance, in Huskies (Figure 6d), while certain EDBUs like EDBU 2 and EDBU 11 show high relative frequency in early developmental stages, the overall distribution tends to broaden as they mature. Other EDBU types, such as EDBU 5 and EDBU 10, also contribute significantly in later age groups, suggesting an expansion of their elemental vocal repertoire. This trend of diversification is also reflected in other breeds. For example, focusing on EDBU1 in German Shepherds (GSD) (Figure 6b), its usage in puppyhood is 24.7%. It peaks at 29.4% during the juvenile period, but then by adulthood, it decreases to 18.1%. Conversely, some EDBUs remain consistently rare as compared to other EDBUs, such as EDBU 1 and EDBU 13 in Shiba Inus (as suggested by Figure 6c, which might indicate their reservation for highly specific contexts or emergence during significant physiological or psychological events, warranting further contextual analysis.

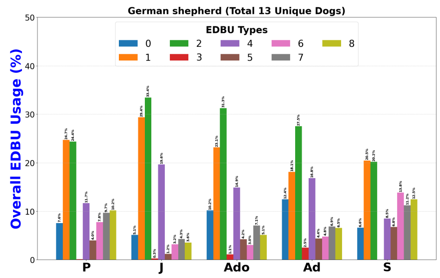
3.4 Bark Sequence Characteristics Over Lifespan

To understand how Barkseqs evolve with age, we characterize Barkseqs with three key metrics. Let B_S denote a single Barkseq, which has an overall start time $T_{start}(B_S)$ and end time $T_{end}(B_S)$.

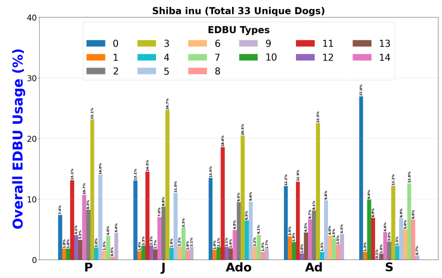
- **Barkseq Overall Duration ($D_{overall_{B_S}}$):** The total time elapsed from its start time ($T_{start}(B_S)$) to its end time ($T_{end}(B_S)$).
- **Barkseq Length (L_{B_S}):** The total count of discrete Bark Units (BUs) constituting the Barkseq B_S.



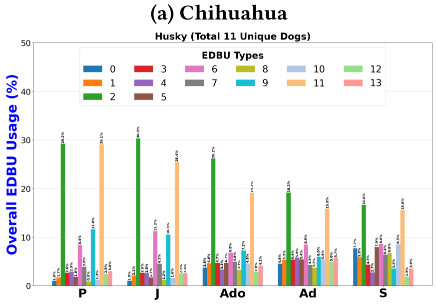
(a) Chihuahua



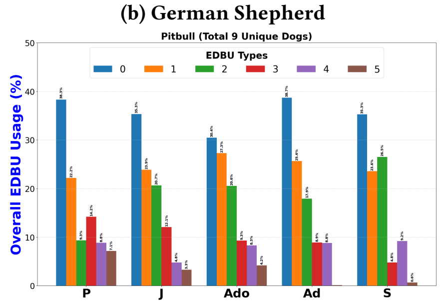
(b) German Shepherd



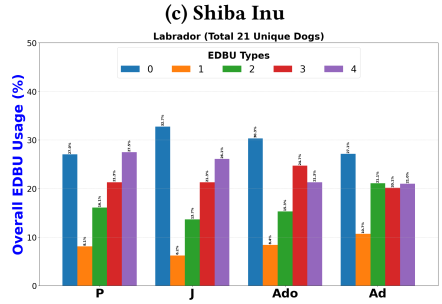
(c) Shiba Inu



(d) Husky

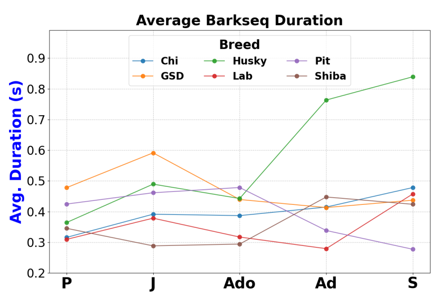


(e) Pitbull

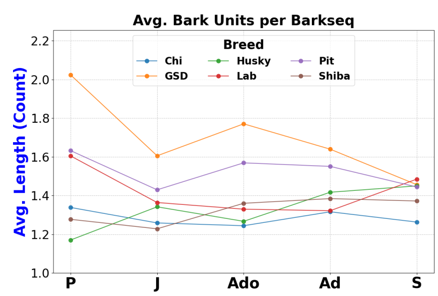


(f) Labrador

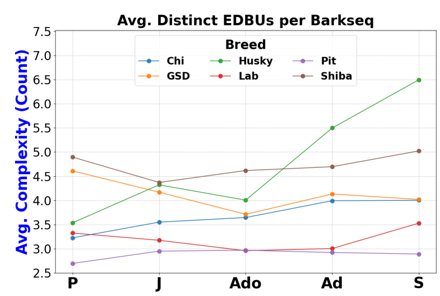
Figure 6: Bar plots showing the micro-averaged percentage distribution of EDBUs by age group for each breed.



(a) Overall Duration.



(b) Length (No. of BUs).



(c) Complexity (Unique EDBUs).

Figure 7: Average Barkseq characteristics (micro average by age group) across breeds: (a) Overall Duration, (b) Length (Number of Bark Units), and (c) Complexity (Number of Unique EDBU types per Barkseq).

- **Barkseq Complexity (C_{B_S}):** The average number of unique Elemental Dog Bark Unit (EDBU) types per Barkseq.

Our analysis of Barkseq characteristics (Figure 7a, Figure 7b, Figure 7c) reveals distinct developmental trajectories that reflect underlying neuromotor maturation, learned vocal strategies, and breed-specific predispositions.

During **puppyhood**, most breeds exhibit relatively short $D_{\text{overall}_{B_S}}$ (e.g., 0.3–0.5s). This brevity may be related to ongoing maturation of vocal production systems. The evolution of vocal apparatus and motor control are general considerations in mammalian vocal development [21], with specific parallels seen in the refinement of motor control in human infant vocalizations [35]. Concurrently, high L_{B_S} values (e.g., 1.4–1.9 BUs per Barkseq) suggest that while sequences are short, they often comprise multiple discrete vocal elements. This pattern of rapid, repetitive vocalizations can be associated with attention-seeking or distress behaviors in dogs [49, 63] during a phase of intense social learning and environmental exploration.

The **juvenile** period marks a critical transition. $D_{\text{overall}_{B_S}}$ generally increases across breeds, indicating improving respiratory control and vocal stamina. Notably, L_{B_S} often decreases for most breeds during this phase. Together, these changes suggest a shift from the rapid, multi-unit vocalizations of puppyhood towards more deliberate and potentially more structured barking patterns. This developmental trajectory likely reflects the maturation of neural circuits influencing vocal production, as neural changes are understood to be crucial for vocal development and learning in other species [28, 55], and aligns with findings on the existence of discoverable phonetic and lexical structures in canine vocalizations [59].

During the **adolescent** phase, breed-specific patterns in $D_{\text{overall}_{B_S}}$ become more pronounced. Some breeds, such as German Shepherds and Labradors, may show continued increases in duration, while others, like Chihuahuas, might exhibit stabilization or slight decreases. This divergence could reflect early manifestations of breed-specific selection pressures on behavior [11], where working or larger breeds might be developing capacities for more sustained vocal communication compared to smaller companion breeds selected for different vocal tendencies.

In **adulthood**, we observe increased inter-breed variability in $D_{\text{overall}_{B_S}}$, with some breeds showing peak durations while others maintain stable patterns established earlier. This heterogeneity likely reflects the full expression of breed-specific vocal phenotypes, where inherent predispositions for vocal behavior, influenced by genetic and developmental factors [20], are now manifested alongside learned, context-dependent vocal responses.

The **senior** period reveals further significant trends for several breeds. Increases in both $D_{\text{overall}_{B_S}}$ and C_{B_S} (particularly evident in Huskies and Labradors, as seen in Figure 7a, Figure 7c) may indicate age-related alterations in vocal control mechanisms. These patterns could be linked to: (1) changes in vocalization patterns associated with aging or cognitive decline, with some studies noting increased vocalization in older dogs [58], (2) compensatory vocal behaviors for reduced vocal efficiency, or (3) pathological changes impacting laryngeal function, such as laryngeal paralysis which is more common in older dogs [42, 57], or changes related to cognitive processing, as Cognitive Dysfunction Syndrome (CDS) has been linked to increased or inappropriate vocalization in aged dogs [43]. The increased complexity (C_{B_S}) observed in some senior dogs might paradoxically represent a degradation of fine vocal control rather than increased sophistication, as aging can affect the precise neural coordination required for consistent bark production.

Throughout all life stages, breed-specific differences in these vocal characteristics likely reflect a combination of functional specialization tied to original breed purposes and artificial selection pressures. For instance, the tendency for sustained high complexity in Huskies (Figure 6d) across age groups may be related to their historical working roles, as selective breeding has shaped unique breed behaviors [11]. In contrast, the relatively more stable and often lower complexity patterns in breeds like Shiba Inus (Figure 6c) may reflect different selection histories.

4 Related Work

Computational methods are increasingly applied to animal communication, a field traditionally in biology, to analyze complex vocal systems and identify their fundamental units, rules, and meanings [5, 7, 46]. Research confirms that canine vocalizations possess structured, language-like properties, with studies linking dog sounds to contextual meaning [61], recognizing emotion [29], categorizing vocalization types [32], and showing that dogs can infer a signaller's size from growls [19]. Building on this, recent computational work has advanced the field by using self-supervised methods to identify phoneme-like units suggesting a rudimentary vocabulary [38, 39], and by developing iterative methods to automatically discover a canine phonetic inventory and its lexical structures [60]. However, computational literature on *age-related changes* in canine vocal patterns is scarce, particularly longitudinal analysis. Prior age-related work includes classifying 8 Mudi dogs into broad age categories (80.25% accuracy) [37] and recent efforts using deep learning on 113 dogs (19,643 barks) [25]. Other datasets include large-scale online data compilations by Wang et al. [59] (>1,300 users, >23 hours) focused on lexical discovery, or Déaux et al. [15] (30 dogs from YouTube), which provided naturalistic data but were smaller or used broad age categories. While valuable, these existing datasets are predominantly cross-sectional. For instance, the study by Pongrácz et al. [49] analyzed barks from Mudi dogs recorded in various situations. Crucially, these resources lack the precise, verified longitudinal tracking of individual dogs needed to model developmental trajectories, highlighting a significant data gap.

Our 'Canine Age Transition Vocalization Dataset' directly addresses this gap. It is a large-scale corpus designed for developmental studies, with 11.4 hours of vocalizations (79,142 Bark Units from 55,718 Barkseqs) from 125 individual dogs across 6 breeds. Sourced from diverse online social media videos, the dataset's core strength lies in its meticulously verified metadata, including precise birthdates, breed, and individual dog IDs, processed through a rigorous pipeline. This careful curation ensures its suitability for robust longitudinal analysis, opening new avenues for future research into vocal development over an individual's lifespan—a dimension previously hindered by data limitations. Furthermore, the introduction of Elemental Dog Bark Units (EDBUs) facilitates a novel, finer-grained analysis of vocal development. The unique synergy of true longitudinal tracking, large scale, and precise metadata enables a new class of developmental studies. For context, automatic age classification from human speech is a well-established research area using machine and deep learning [36, 50, 54, 56], underscoring the potential for similar advances in canine research.

5 Conclusion

This paper presents a rare dataset and novel findings on canine vocal pattern development, which lays a critical foundation for understanding canine vocal evolution and opens several avenues for future work. A key next step is the detailed acoustic characterization of EDBUs (e.g., formants, pitch contours, energy distributions) per type, within and across breeds, potentially establishing a comprehensive acoustic 'dictionary' of these units. Expanding the dataset, particularly for underrepresented breed/age cohorts, will also enhance statistical power and generalizability. Future research could also explore the semantic content or contextual correlates of EDBUs and their sequences, moving beyond distributional analysis to investigate functional meaning. Examining the interplay of innate maturational trajectories with environmental or learning influences on vocal development, including cross-breed EDBU comparisons, presents another promising direction. Furthermore, correlating observed vocal changes with known milestones in canine cognitive or physical development could offer a more holistic view of aging in dogs. By facilitating such multifaceted research, this dataset is poised to significantly advance the field, providing a more nuanced perspective on lifelong vocal communication in canines.

Acknowledgments

This work was partially supported by NSF Award No. 2349713. Our gratitude goes to Mengyue Wu for her useful discussions.

References

[1] Francesc Alías, Joan Claudi Socoró, and Xavier Sevillano. 2016. A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences* 6, 5 (2016), 143.

[2] American Kennel Club. 2025. Chihuahua Dog Breed Information — akc.org. https://www.akc.org/dog-breeds/chihuahua/. [Accessed 18-05-2025].

[3] American Kennel Club. 2025. Shiba Inu Dog Breed Information — akc.org. https://www.akc.org/dog-breeds/shiba-inu/. [Accessed 18-05-2025].

[4] American Kennel Club. 2025. Siberian Husky Dog Breed Information — akc.org. https://www.akc.org/dog-breeds/siberian-husky/. [Accessed 18-05-2025].

[5] Jacob Andreas, Gašper Beguš, Michael M Bronstein, Roee Diamant, Denley Delaney, Shane Gero, Shafi Goldwasser, David F Gruber, Sarah de Haas, Peter Malkin, et al. 2022. Toward understanding the communication in sperm whales. *Iscience* 25, 6 (2022).

[6] Bella+Duke. 2023. Labrador owner's dog breed guide - Bella+Duke — bellaandduke.com. https://www.bellaandduke.com/dogs/expert-advice/dog-breeds/labrador-owners-guide/. [Accessed 18-05-2025].

[7] Christian Bergler, Manuel Schmitt, Rachael Xi Cheng, Andreas K Maier, Volker Barth, and Elmar Nöth. 2019. Deep Learning for Orca Call Type Identification-A Fully Unsupervised Approach.. In *Interspeech*. 3357–3361.

[8] Christopher M Bishop and Nasser M Nasrabadi. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer.

[9] Sanyuan Chen, Yu Wu, Chengyi Wang, Shujie Liu, Daniel Tompkins, Zhuo Chen, and Furu Wei. 2022. Beats: Audio pre-training with acoustic tokenizers. *arXiv preprint arXiv:2212.09058* (2022).

[10] D. Caroline Coile. 2021. How to Help Pit Bulls Age Well | dummies — dummies.com. https://www.dummies.com/article/home-auto-hobbies/pets/dogs/breeds/pit-bulls/how-to-help-pit-bulls-age-well-275334/. [Accessed 18-05-2025].

[11] Raymond Coppinger and Lorna Coppinger. 2001. *Dogs: A startling new understanding of canine origin, behavior & evolution.* Simon and Schuster.

[12] Nicole Cosgrove. 2025. How Long Do German Shepherds Live? Average Lifespan, Data & Care Guide - Dogster — dogster.com. https://www.dogster.com/dog-health-care/german-shepherd-lifespan-how-long-do-they-live. [Accessed 18-05-2025].

[13] David L Davies and Donald W Bouldin. 2009. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence* 2 (2009), 224–227.

[14] Steven Davis and Paul Mermelstein. 1980. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing* 28, 4 (1980), 357–366.

[15] Eloïse C Déaux, Théophane Piette, Florence Gaunet, Thierry Legou, Luc Arnal, and Anne-Lise Giraud. 2024. Dog–human vocal interactions match dogs' sensory-motor tuning. *PLoS biology* 22, 10 (2024), e3002789.

[16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 4171–4186.

[17] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[18] Georgios Evangelopoulos and Petros Maragos. 2006. Multiband modulation energy tracking for noisy speech detection. *IEEE Transactions on audio, speech, and language processing* 14, 6 (2006), 2024–2038.

[19] Tamás Faragó, Peter Pongracz, Ádám Miklósi, Ludwig Huber, Zsófia Virányi, and Friederike Range. 2010. Dogs' expectation about signalers' body size by virtue of their growls. *PLoS One* 5, 12 (2010), e15175.

[20] Tamás Faragó, Simon Townsend, and Friederike Range. 2013. The information content of wolf (and dog) social communication. In *Biocommunication of animals*. Springer, 41–62.

[21] W Tecumseh Fitch. 2000. The evolution of speech: a comparative review. *Trends in cognitive sciences* 4, 7 (2000), 258–267.

[22] Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 776–780.

[23] Shane Gero, Hal Whitehead, and Luke Rendell. 2016. Individual, unit and vocal clan level identity cues in sperm whale codas. *Royal Society Open Science* 3, 1 (2016), 150372.

[24] Theodoros Giannakopoulos and Aggelos Pikrakis. 2014. *Introduction to audio analysis: a MATLAB® approach.* Academic Press.

[25] José Ramón Gómez-Armenta, Humberto Pérez-Espinosa, José Alberto Fernández-Zepeda, and Verónica Reyes-Meza. 2024. Automatic classification of dog barking using deep learning. *Behavioural Processes* 218 (2024), 105028.

[26] Yuan Gong, Yu-An Chung, and James Glass. 2021. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778* (2021).

[27] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783* (2024).

[28] Kurt Hammerschmidt and Julia Fischer. 2008. Constraints in primate vocal production. *Evolution of communicative* (2008), 93–119.

[29] Simone Hantke, Nicholas Cummins, and Bjorn Schuller. 2018. What is my dog trying to tell me? The automatic recognition of the context and perceived emotion of dog barks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5134–5138.

[30] Naomi D Harvey. 2021. How old is my dog? Identification of rational age groupings in pet dogs based upon normative age-linked processes. *Frontiers in veterinary science* 8 (2021), 643085.

[31] Jessica M Hoffman, Dan G O'Neill, Kate E Creevy, and Steven N Austad. 2018. Do female dogs age differently than male dogs? *The Journals of Gerontology: Series A* 73, 2 (2018), 150–156.

[32] Jieyi Huang, Chunhao Zhang, Mengyue Wu, and Kenny Zhu. 2023. Transcribing vocal communications of domestic shiba lnu dogs. In *Findings of the Association for Computational Linguistics: ACL 2023*. 13819–13832.

[33] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, and Raj Reddy. 2001. *Spoken Language Processing: A guide to theory, algorithm, and system development.* Prentice hall PTR.

[34] Ana G. Jiménez. 2024. A revisiting of 'the hallmarks of aging' in domestic dogs: current status of the literature. *GeroScience* 46 (Feb. 2024), 241–255. doi:10.1007/s11357-023-00911-5

[35] Raymond D Kent and Ann D Murray. 1982. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America* 72, 2 (1982), 353–365.

[36] Damian Kwasny and Daria Hemmerling. 2021. Gender and age estimation methods based on speech using deep neural networks. *Sensors* 21, 14 (2021), 4785.

[37] Ana Larranaga, Concha Bielza, Péter Pongrácz, Tamás Faragó, Anna Bálint, and Pedro Larranaga. 2015. Comparing supervised learning methods for classifying sex, age, context and individual Mudi dogs from barking. *Animal cognition* 18, 2 (2015), 405–421.

[38] Xingyuan Li, Sinong Wang, Zeyu Xie, Mengyue Wu, and Kenny Q Zhu. 2024. Phonetic and lexical discovery of a canine language using hubert. *arXiv preprint arXiv:2402.15985* (2024).

[39] Xingyuan Li, Kenny Zhu, and Mengyue Wu. 2025. Dog2vec: Self-Supervised Pre-Training for Canine Vocal Representation. In *Proceedings of the 26th Interspeech Conference.*

[40] Xubo Liu, Haohe Liu, Qiuqiang Kong, Xinhao Mei, Jinzheng Zhao, Qiushi Huang, Mark D Plumbley, and Wenwu Wang. 2022. Separate what you describe: Language-queried audio source separation. *arXiv preprint arXiv:2203.15147* (2022).

[41] Jamie Lovejoy. 2024. How Long Do Dogs Live? https://www.petmd.com/dog/care/how-long-do-dogs-live. Updated June 13, 2024. Reviewed by Kelly Gold, DVM. Accessed: 2025-05-18.

[42] Catriona M MacPhail and Eric Monnet. 2001. Outcome of and postoperative complications in dogs undergoing surgical treatment of laryngeal paralysis: 140 cases (1985–1998). *Journal of the American Veterinary Medical Association* 218, 12 (2001), 1949–1956.

[43] Brennen A McKenzie, Frances Chen, and Michael L LaCroix-Fralish. 2022. The phenotype of aging in the dog: how aging impacts the health and well-being of dogs and their caregivers. *Journal of the American Veterinary Medical Association* 260, 9 (2022), 963–970.

[44] Kellyn McNulty. 2020. Beyond gray muzzles: defining aging in dogs — Dog Aging Project. https://dogagingproject.org/beyond-gray-muzzles-defining-aging-in-dogs. [Accessed 18-05-2025].

[45] Yunbi Nam, Michelle White, Elinor K Karlsson, Kate E Creevy, Daniel EL Promislow, Robyn L McClelland, and Dog Aging Project Consortium. 2024. Dog size and patterns of disease history across the canine age spectrum: Results from the Dog Aging Project. *PLoS One* 19, 1 (2024), e0295840.

[46] Tuomas Oikarinen, Karthik Srinivasan, Olivia Meisner, Julia B Hyman, Shivangi Parmar, Adrian Fanucci-Kiss, Robert Desimone, Rogier Landman, and Guoping Feng. 2019. Deep convolutional network for animal sound classification and source attribution using dual audio recordings. *The Journal of the Acoustical Society of America* 145, 2 (2019), 654–662.

[47] Mimi Padmabandu. 2022. How to Care for Your Dog at Every Life Stage — embarkvet.com. https://embarkvet.com/resources/dog-care-every-age/. [Reviewed by Jenna Dockweiler, MS, DVM, DACT, CCRT, CVAT. Accessed 18-05-2025].

[48] Michael A Pardo, David S Lolchuragi, Joyce Poole, Petter Granli, Cynthia Moss, Iain Douglas-Hamilton, and George Wittemyer. 2024. Female African elephant rumbles differ between populations and sympatric social groups. *Royal Society Open Science* 11, 9 (2024), 241264.

[49] Péter Pongrácz, Csaba Molnár, Adám Miklósi, and Vilmos Csányi. 2005. Human listeners are able to classify dog (Canis familiaris) barks recorded in different situations. *Journal of comparative psychology* 119, 2 (2005), 136.

[50] Zakariya Qawaqneh, Arafat Abu Mallouh, and Buket D Barkana. 2017. Deep neural network framework and transformed MFCCs for speaker's age and gender classification. *Knowledge-Based Systems* 115 (2017), 5–14.

[51] Lawrence R Rabiner. 1978. *Digital processing of speech signals*. Pearson Education India.

[52] Lawrence R Rabiner and Marvin R Sambur. 1975. An algorithm for determining the endpoints of isolated utterances. *Bell System Technical Journal* 54, 2 (1975), 297–315.

[53] Dzati Athiar Ramli and Haryati Jaafar. 2016. Peak finding algorithm to improve syllable segmentation for noisy bioacoustic sound signal. *Procedia Computer Science* 96 (2016), 100–109.

[54] Shijing Si, Jianzong Wang, Junqing Peng, and Jing Xiao. 2022. Towards speaker age estimation with label distribution learning. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4618–4622.

[55] Charles T Snowdon. 2017. Learning from monkey 'talk'. *Science* 355, 6330 (2017), 1120–1122.

[56] Werner Spiegl, Georg Stemmer, Eva Lasarcyk, Varada Kolhatkar, Andrew Cassidy, Blaise Potard, Stephen Shum, Young Chol Song, Puyang Xu, Peter Beyerlein, et al. 2009. Analyzing features for automatic age estimation on cross-sectional data.. In *Interspeech*, Vol. 2009. 2923.

[57] Kelley M Thieman, DJ Krahwinkel, Michael H Sims, and G Diane Shelton. 2010. Histopathological confirmation of polyneuropathy in 11 dogs with laryngeal paralysis. *Journal of the American Animal Hospital Association* 46, 3 (2010), 161–167.

[58] Lisa J. Wallis, Dóra Szabó, Boglárka Erdélyi-Belle, and Enikö Kubinyi. 2018. Demographic Change Across the Lifespan of Pet Dogs and Their Impact on Health Status. *Frontiers in Veterinary Science* 5 (2018). doi:10.3389/fvets.2018.00200

[59] Theron Wang, Xingyuan Li, Chunhao Zhang, Mengyue Wu, and Kenny Zhu. 2024. Phonetic and Lexical Discovery of Canine Vocalization. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 13972–13983.

[60] Theron S Wang, Xingyuan Li, Hridayesh Lekhak, Tuan Minh Dang, Mengyue Wu, and Kenny Zhu. 2025. Toward Automatic Discovery of a Canine Phonetic Alphabet. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 9207–9219.

[61] Yufei Wang, Chunhao Zhang, Jieyi Huang, Mengyue Wu, and Kenny Zhu. 2023. Towards lexical analysis of dog vocalizations via online videos. *arXiv preprint arXiv:2309.13086* (2023).

[62] Marina M Watowich, Evan L MacLean, Brian Hare, Josep Call, Juliane Kaminski, Ádám Miklósi, and Noah Snyder-Mackler. 2020. Age influences domestic dog cognitive performance independent of average breed lifespan. *Animal cognition* 23 (2020), 795–805.

[63] Sophia Yin and Brenda McCowan. 2004. Barking in domestic dogs: context specificity and individual identification. *Animal behaviour* 68, 2 (2004), 343–355.