OPEN ACCESS
VOL. 5, NO. 2, PP.144-154, MAY 2023

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/INISTA.V5I2.1091

Journal of Informatics, Information System, Software Engineering and Applications (INISTA)

Automatic Birdsong Splitting and Syllabic Analysis of Jalak Suren

Agi Prasetiadi *[1], Julian Saputra [2]

Faculty of Informatics, Institut Teknologi Telkom Purwokerto
Jl. DI Panjaitan No. 128, Purwokerto, 53147, Indonesia

[*1] agi@ittelkom-pwt.ac.id
[2] 19102008@ittelkom-pwt.ac.id

Received on 01-06-2023, revised on 20-06-2023, accepted on 10-07-2023

Abstract

The study of birdsong has received relatively limited attention in the field of artificial intelligence, despite its long-standing intrigue and the question of whether birds possess a form of language. Previous research has provided evidence suggesting the presence of structurally organized words recognized by birds, such as the strong reactions observed in Japanese tits and Pied babblers when exposed to specific sequences of artificially played calls. Altering the speed of a sequence also influences the birds' responses, further supporting the existence of organized linguistic units in avian vocalizations. In this study, we propose a novel approach for analyzing birdsong by employing automatic syllable segmentation and syllabic similarity analysis. Our focus is on the Jalak Suren species (Sturnus contra), renowned for its melodious song. Through the identification and categorization of distinct syllabic units in birdsong recordings, we investigate the statistical occurrence of these syllables within the sequence of birdsong. Our findings reveal remarkable similarities between the statistical occurrence of syllables in birdsong and those found in human language passages.

Keywords: Birdsong, Jalak Suren, Automatic syllable segmentation, Syllabic analysis.

Corresponding Author:
*Agi Prasetiadi
Faculty of Informatics, Institut Teknologi Telkom Purwokerto
Jl. DI Panjaitan No. 128, Purwokerto, Central Java, 53147, Indonesia
Email: agi@ittelkom-pwt.ac.id

I. INTRODUCTION

BIRDSONG serves various functions, including territorial defense, attracting mates, and individual recognition. It is a primary means of communication between birds of the same species and can convey information about the bird's identity, fitness, and reproductive status [1]. Birds display distinct vocal variations known as dialects, which differ among different populations or regions [2] [3].

Japanese tits combine different calls in ordered sequences to convey compound messages, with specific orderings carrying distinct meanings. Altering the speed of a sequence affects the bird's response [4]. Pied babblers also combine distinct vocalizations into larger sequences, with the function of the sequence related to its constituent parts. Experiments suggest babblers process the sequence compositionally, with mobbing sequences eliciting strong responses compared to individual calls or control sequences [5].

Thus similar to human, birds may rely on two main key components of language, that are sound and grammar [6] [7]. A syllable is a basic unit that structures speech sounds, consisting of a nucleus (usually a vowel) and optional margins (typically consonants). Syllables serve as foundational components in word construction and organization at the phonological level [8]. Languages vary in phoneme usage, with some like Hawaiian showing repetitive phonemes due to limited inventory, while others like Taa exhibit diverse phoneme variations due to a larger inventory. In language, cognitive reference can take three primary forms:

iconic, indexical, and symbolic [9]. In the context of bird communication, we can consider the chirp as indexical reference, or as the equivalent of a syllable within birdsong.

In this research, we employ an automated approach to segment the birdsong of Jalak Suren into distinct syllables and subsequently examine their syllabic characteristics. By extracting a distinct inventory of syllables, we establish symbolic representations for each syllable, enabling us to interpret the birdsong as a conventional textual sequence. If the sequence exhibits language-like properties, it is expected to demonstrate similar patterns of word entropy observed in other languages [10]. Jalak Suren (Sturnidae family) is a widely recognized avian species in Indonesia renowned for its melodic and repetitive vocalizations [11]. Figure 1 illustrates the consistent popularity of Jalak Suren among the Indonesian population [12]. Lovebird initially dominated Google search trends with 44.3% presence in May 2018, but its dominance decreased over time to only 12% in 2023. Conversely, Jalak initially held a 10.4% dominance, but it experienced consistent growth and reached 16% in 2023.

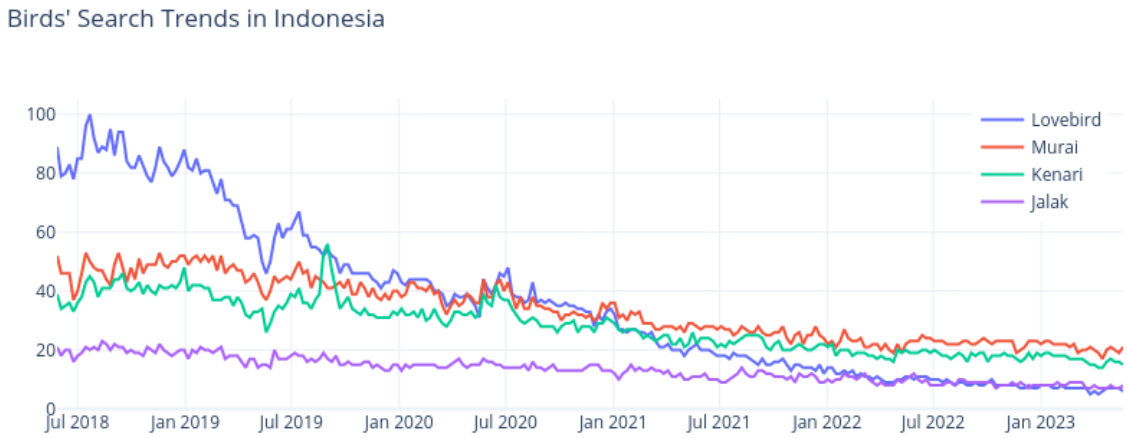


Fig. 1. Search trends related to birds in Indonesia over the previous five-year period according to Google Trends.

The research is structured as follows: First, we will apply techniques such as silent interval analysis, masking, and syllable extraction. Next, we will explore the similarity among the syllables by utilizing special syllable categorization algorithms, aiming to group them into a set of unique syllables based on their length and similarity. Subsequently, we will examine the patterns of similarity and repetition within the identified syllables to determine if they are purposeful or random in nature. Finally, the syllables will be converted into a symbolic sequence, and their word occurrence frequency will be calculated to assess their linguistic properties.

II. RESEARCH METHOD

A. *Data Collection*

Two types of data were collected for this study. The first dataset was obtained from YouTube recordings capturing the vocalizations of Jalak Suren birds during their morning singing sessions. The audio recording were specifically chosen to ensure a duration of 4 minutes. With a sampling rate of 44100 data points per second, this recordings provide a quite detailed representation of the birdsong patterns.

The second dataset involves passages from young individuals undergoing a trial for a death sentence, simulating an environment comparable to the conditions experienced by the analyzed Jalak Suren bird, which is also confined within a birdcage. By including this similar dataset, the study seeks to explore potential similarity in syllabic elements of human and bird vocalizations under such circumstances.

B. *Silent Interval*

In order to facilitate the segmentation of syllable within birdsong, it is necessary to identify and mark the positions of potential silent intervals. This can be achieved by utilizing a straightforward standard deviation calculation. By employing standard deviation, the detection of silent intervals is enhanced, as it

tends to reduce the amplitude of silent intervals to values close to zero while amplifying other sounds, thereby rendering their positions more visible within the entire birdsong sequence.

$$s_i = \sqrt{\frac{\sum_{j=i}^{i+w}(x_j - \mu_i)^2}{w-1}} \tag{1}$$

The transformation of raw audio data into a sequence of windowed standard deviation is outlined in Equation (1). Within this equation, μ_i represents the average value of the audio data within a designated window time, computed as $\Sigma_{j=i}^{i+w} x_j/w$. This formulation enables the calculation of the standard deviation by evaluating the audio data in consecutive windows, where each window spans from index i to $i+w$.

C. Masking

Subsequently, the amplified silent interval signal is subjected to quantization using a threshold to obtain a binary signal, denoted as u. In this study, a threshold value of *1200* (designated as τ) was utilized, although this value is arbitrary and can be adjusted according to the specific circumstances. The quantized silent interval is computed using the formula $u = \text{threshold}(s, \tau) = (s > \tau) \times 1$. Following this, the quantized silent intervals are prepared according to Equation (2), ensuring a smooth outcome for the subsequent processes, where the value of M is set to *1000*.

$$u' = u_{M-1:1:-1} \parallel u \parallel u_{1:M-1} \tag{2}$$

Once the matrix u' is prepared, the next step involves convolving all quantized silent intervals, with a Hann window, denoted as h. Equation (3) defines the Hanning window, where $0 \leq n \leq M-1$. The selection of the Hanning window is motivated by its ability to provide a smoother transient effect [13], effectively addressing the initial and final portions of the syllables.

$$h_n = 0.5 - 0.5\cos\left(\frac{2\pi n}{M-1}\right) \tag{3}$$

Since there will be presence of occasional cracks, to obtain the smooth mask, m, the Equation (4) can be employed. This process will minimaze the cracks into small bumps that can be ignored by further quantization on mask m.

$$m = h * u' \tag{4}$$

The clear cut syllables, denoted as X, can be generated by performing the bitwise multiplication operation between the original sound signal, represented as x, and the smooth mask, denoted as m, as described in Equation (5).

$$X = x \odot m \tag{5}$$

D. Syllable Extraction

During this stage, our objective is to extract all the syllables present within the birdsong data and compile them into a syllable list. It is important to note that the syllables within the list are not necessarily unique, similar to a collection of words within a document. Algorithm 1 outlines the process of sequentially extracting masked syllables from the data and appending them to the syllable list.

Audio data differs from text data in that there is no clear separation between words. Additionally, determining the minimum duration of a chirp that is still recognizable by birds is a challenge. In our algorithm, we define the minimum limit of a chirp recognized as a single syllable as ε, which we set to *50* or *1.13* ms.

E. Syllabic Categorization

A similarity analysis of all syllables will be conducted to cluster similar chirps into distinct groups or categories, ultimately establishing a set of unique syllables. Given the flexibility of syllable signals and the potential variations in shape even when the sounds appear similar, it is crucial to apply specific normalization techniques to minimize shape differences. The initial step involves applying a low-pass filter to each syllable.

To achieve signal smoothing, we employ the Discrete Cosine Transform (DCT) on the syllable signals [14]. This transformation technique, similar to the Fourier Transform [15], converts the signals from the

time domain to the frequency domain. However, unlike the Fourier Transform, the DCT provides real-valued results. In this research, we utilize DCT type 2, as shown in Equation (6), where the original signal is denoted as x and the transformed signal in the frequency domain is represented as y. A low-pass filter with a cutoff frequency of ε Hz is applied during this process. This smoothing approach differs from the silent interval technique, which employs standard deviation to assess signal volatility within specific time windows. By bypassing high-frequency components, the smoothed signal provides an approximation of the area associated with a specific syllable.

$$y_k = 2\sum_{n=0}^{N-1} x_n \cos\left(\frac{\pi k(2n+1)}{2N}\right) \tag{6}$$

The smoothing process ensures that even if there are slight variations in the original syllable signal, the resulting smoothed signal remains similar. This property makes the DCT-based smoothed signal a favorable choice for assessing signal similarity. The detail of pseudocode can be seen at Algorithm 2.

Algorithm 1 SyllableExtractor: This algorithm takes the masked syllables, denoted as input X, and produces a list of syllables, represented as output p, in a list format.

```
function SyllableExtractor(X)
    p ← {}                                          ▷ p for syllables
    (f, l, s) ← (0, 0, 0)                           ▷ first note, last note, state
    if X_0 = 0 then                                 ▷ State of first mask element
        s ← 0
    else
        s ← 1
    for i ← 0 to |X| − 1 do                         ▷ Scan all data
        if s = 0 then
            if X_{i+1} ≠ 0 then                     ▷ Found new syllable
                (f, s) ← (i + 1, 1)
        else
            if X_{i+1} = 0 then
                (l, s) ← (i + 1, 0)
                if l − f > ε then
                    p ← p ∪ X_{f:l}                 ▷ End of a syllable
    → p
```

Algorithm 2 LPF: Smooth extractor based on the DCT low-pass filter.

```
function LPF(f)
    f' ← threshold(f, max(f)/2)                     ▷ Convert to binary signal
    f' ← iDCT(DCT(f')_{:ε} ‖ O_{1×|f|−ε})
    → f'
```

The Jaccard similarity coefficient is utilized to assess the similarity between syllables in the next step. Equation (7) outlines the calculation method for determining the similarity between two sets [16]. The approach involves comparing the intersection and union of the two regions. A Jaccard similarity coefficient of 1 indicates that the two regions have identical shapes and area sizes. Conversely, a coefficient of 0 indicates no similarity between the two regions.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

Figure 2 illustrates the comparison between two regions. The left image represents the gray area formed by the intersection of the sound graphs from the 1st and 2nd sounds, while the right image displays the green area formed by the union of the sound graphs from the 1st and 2nd sounds.

Algorithm 3 demonstrates the computation and storage of Jaccard Similarity Coefficients in matrix S. Initially, all syllables undergo smoothing using Algorithm 2. Subsequently, each syllable is paired with every other syllable. If the difference in their lengths exceeds the threshold ϵ_l, the calculation is disregarded

and their similarity is set to 0. In our specific case, ϵ_l is defined as ε. Conversely, if the length difference is within the threshold, the similarity between the syllables is computed and recorded in matrix *S*.

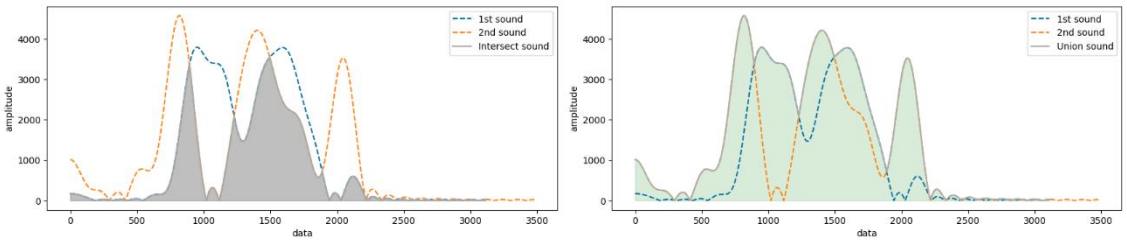


Fig. 2. Intersection and Union Sound Graph Comparison.

Algorithm 3 Syllable Similarity Scanning Algorithm: Constructing Similarity Matrix.

$S \leftarrow O_{|p|\times|p|}$ ▷ Zeros matrix
$p' \leftarrow \{\}$
for $i \leftarrow 0$ to $|p| - 1$ **do** ▷ Smooth all syllables
  $p' \leftarrow p' \cup |\text{LPF}(p_i)|$
for $i \leftarrow 0$ to $|p| - 1$ **do** ▷ Scan all syllables
  $l_s \leftarrow |p_i|$ ▷ l_s as length of source syllable
  for $j \leftarrow i + 1$ to $|p|$ **do**
    $l_t \leftarrow |p_i|$ ▷ l_t as length of target syllable
    if $\frac{|l_t - l_s|}{l_s} < \epsilon_l$ **then** ▷ Lengths must be similar
      $S_{j,i} \leftarrow S_{i,j} \leftarrow J(p'_i, p'_j)$

Algorithm 4 is designed to group syllables into unique sets based on their similarity. It utilizes several variables: *t* represents the collection of all possible unique syllable groups, *v* is a list that keeps track of the indices of syllables that have been checked for similarity, *p* is the list of input syllables, which may contain duplicates, *k* represents the indices of syllables that exhibit a high Jaccard similarity coefficient (stored in the similarity matrix *S*) with the *i*-th syllable. If the length of *k* is zero, it implies that the syllable is unique and does not have any similarity with other syllables. *n* is used to determine if a syllable is new and not yet included in any existing group. It counts the number of similar syllables that have not been visited (i.e., not in *v*). In our study, we have defined the value of ϵ_s as 0.55, indicating that we consider a syllable to be the same as the targetted syllable if it exhibits a minimum similarity of 55%.

First, it creates a set *o* containing only the current syllable at index *i*. Then, it identifies the indices of syllables (*k*) that have a similarity coefficient above the threshold (ϵ_s) with the syllable at index *i*. If the size of *k* is zero, it means the syllable is unique (has no similar syllables). During the algorithm's execution, if similar syllables have already been added to the visited variable *v*, the algorithm traces back to identify the corresponding syllable group and adds itself to that group. However, if the syllable is indeed new, it proceeds to add each of its similar syllables one by one into its newly created group. Ultimately, at the conclusion of the algorithm, variable *t* represents the list of unique syllable groups.

F. Syllables Statistic

In the final phase of the analysis, we will compare the behavior of birdsong syllables with those found in a passage from the Indonesian language that has a similar duration. The selected Indonesian passage, consisting of approximately 355 words, depicts the inner thoughts of a young boy who has endured prolonged imprisonment and desires liberation. This particular passage was chosen to simulate a comparable environment to that experienced by the analyzed Jalak Suren bird, which is also confined within

a birdcage. The examination of the Indonesian passage reveals the presence of 874 syllables, of which 251 are unique.

Algorithm 4 Syllable Grouping Algorithm for generating unique set of syllables.

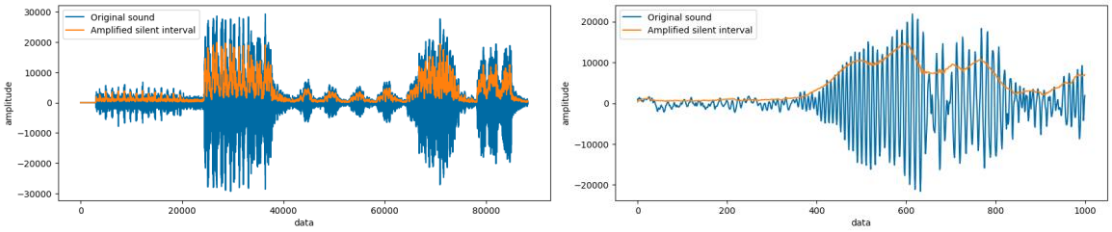
$(t, v) \leftarrow (\{\}, \{\})$
for $i \leftarrow 0$ to $|p| - 1$ **do**
  if $i \notin v$ **then**
    $o \leftarrow \{i\}$
    $v \leftarrow v \cup \text{set}(o)$
    $k \leftarrow \{j \in [0, |p| - 1] : S_{i,j} > \epsilon_s\}$
    if $|k| = 0$ **then** ▷ Unique syllable
      continue
    $n \leftarrow 0$ ▷ Is this new syllable?
    for $j \in k$ **do**
      $n \leftarrow n + (j \notin v) * 1$
    if $n = 0$ **then** ▷ Add into already registered list
      for $j \leftarrow 0$ to $|t| - 1$ **do**
        if $k_0 \in t_j$ **then**
          $t_j \leftarrow t_j \cup \text{set}(o)$
          break
      continue
    for $j \in k$ **do** ▷ Collect similar syllable(s)
      if $j \notin v$ **then**
        $o \leftarrow o \cup \text{set}(j)$
        $v \leftarrow v \cup \text{set}(j)$
    $t \leftarrow t \cup \text{set}(o)$ ▷ t as unique syllable groups

III. RESULTS AND DISCUSSION

A. *Silent Interval*

Figure 3.a illustrates the unprocessed birdsong data captured within a two-second interval, indicated by the blue line. This data is denoted as *x* and serves as the basis for analysis. The orange highlighted section represents the specific region used for detecting the silent interval, denoted as *s*. Notably, all values within the amplified silent interval signal, *s*, are above *0*, providing an advantageous characteristic for the development of a local filter.

Figure 3.b provides a visual representation of the zoomed amplified silent interval. The observed portion exhibits a smooth unipolar signal, where higher values indicate increased variance or volatility during that specific time interval. In this study, a window size, *w*, of 40 was employed.



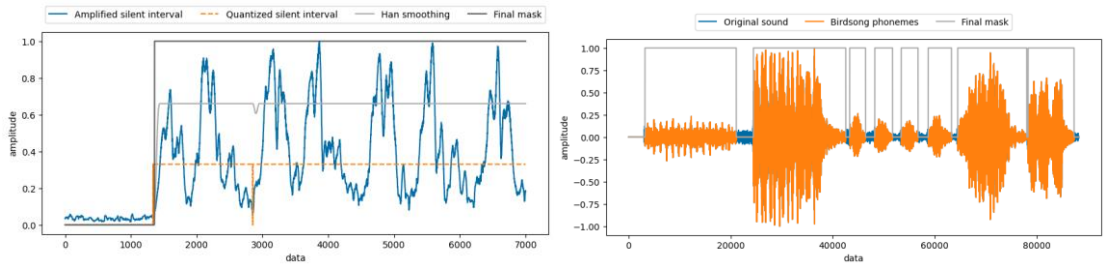
(a) (b)

Fig. 3. The original birdsong signal (blue) and the amplified silent interval (orange): a) The sound is extracted from the first 2 seconds; and b). the zoomed-in view of the original birdsong and the amplified silent interval extracted from sequence data 24000-th to 25000-th of the Jalak Suren birdsong.

B. *Masking*

Next, we create mask step by step using Hanning window. By referring to Figure 4.a, the transformation of the amplified silent interval into the quantized silent interval can be observed, revealing the presence of

occasional cracks. These cracks will be minimized by convoluting the signal with Hann window. Finally the final mask is obtained by thresholding the convoluted mask.



(a) (b)

Fig. 4. Syllable Segmentation Process: a) Refinement Process for Syllable Mask, involving quantized silent intervals, Hann smoothing, and the final mask; and b) Segmentation of Birdsongs into Syllables.

Figure 4.b illustrates the final phase of applying the final mask to the original birdsong, resulting in the segmentation of the audio into discernible syllables. This type of signal offers an advantage due to its simplicity in detecting regions with sound (non-zero values) and regions without sound (zero values).

C. Syllable Extraction

After that, we run syllable extraction on masked signal. Upon running the algorithm, we identified a total of 757 syllables from a 4-minute recording of Jalak Suren's babbling. It is worth noting that there were a few syllables detected within the interval of *14* to *40* or *0.3* ms to *0.9* ms, but due to their small number, they were discarded by the algorithm. The smallest syllables captured by this algorithm had a duration of approximately *200* or *4.5* ms.

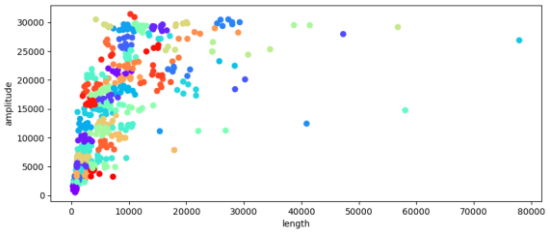


Fig. 5. Scatter plot of recorded syllables, depicting their length and maximum amplitude. Syllables are grouped into 16 clusters using the KMeans algorithm.

Figure 5 presents a scatter plot of recorded syllables, showcasing their length and maximum amplitude as key features. This features are then clustered using KMeans algorithm. However, utilizing these features alone to group syllables into unique categories using clustering algorithms proves to be unsuitable. In this particular example, we employed the algorithm with 64 clusters to cluster the syllables. It should be noted that the number of unique syllables is still unknown, and it could be any value, not necessarily 64. Despite running the silhouette score, the best score suggests that the syllable collection should be grouped into only

2 or 3 categories, which is incorrect. Hence, the development of a specialized algorithm for syllabic categorization is necessary.

D. Syllabic Categorization

In the first step of syllabic categorization, the signal needs to be smoothed. Figure 6 presents the original syllable signal (blue line) and the corresponding smoothed signal obtained by applying a low-pass filter based on the Discrete Cosine Transform (DCT) (orange line).

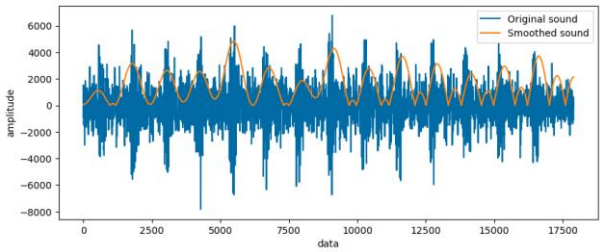


Fig. 6. Comparison of the original birdsong signal and the low-pass filtered signal smoothed using Discrete Cosine Transform (DCT).

Subsequently, the similarity between all possible pairs of phonemes should be computed, and the resulting similarity values should be stored in the similarity matrix. The similarity matrix, as depicted in Figure 7, provides a visual representation of the calculated similarities between phoneme pairs. Interestingly, the repetitive nature of certain groups can be observed from the matrix. Specifically, the syllables in the region encompassing the 8th to 28th positions demonstrate frequent consecutive repetitions. This observation reinforces the notion that individual bird chirps do not necessarily represent specific meanings on their own, as they serve as the fundamental syllables within a broader concept, namely words. But it is worth to note that since the syllables are repetitive, it may indicate structured word or grammar within birdsong.

Following that, the Syllable Grouping Algorithm is applied to identify a list of distinct syllables. As the result, a total of 353 syllables were categorized into 77 distinct syllable groups. Additionally, 404 unique syllables were identified without any repetition in the dataset, resulting in a total of 481 unique syllables within the birdsong data.

Figure 8 presents a visual representation of 9 syllables that have been assigned as unique groups. The first 6 plots exhibit accurately grouped syllables, demonstrating distinct clusters of syllable patterns. However, the last 3 plots depict less precise grouping, although some resemblance between the syllables can still be observed. Notably, the 6th plot reveals an intriguing pattern, indicating that Jalak Suren

consistently produces exact or highly similar syllables within a short time period. This finding suggests that the birdsong of Jalak Suren is not random but follows a structured pattern.

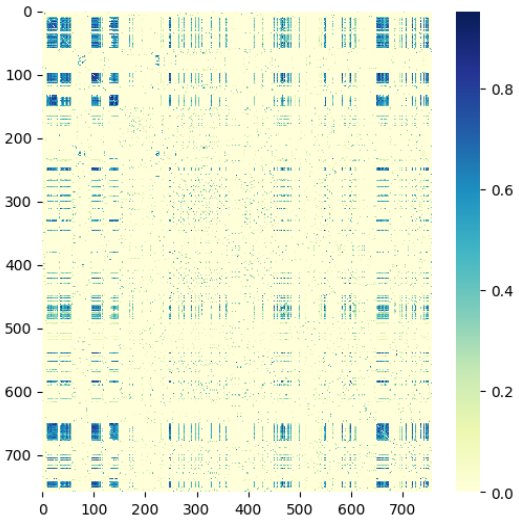


Fig. 7. Similarity matrix based on Jaccard Similarity Coefficients across all possible pairs of syllables.

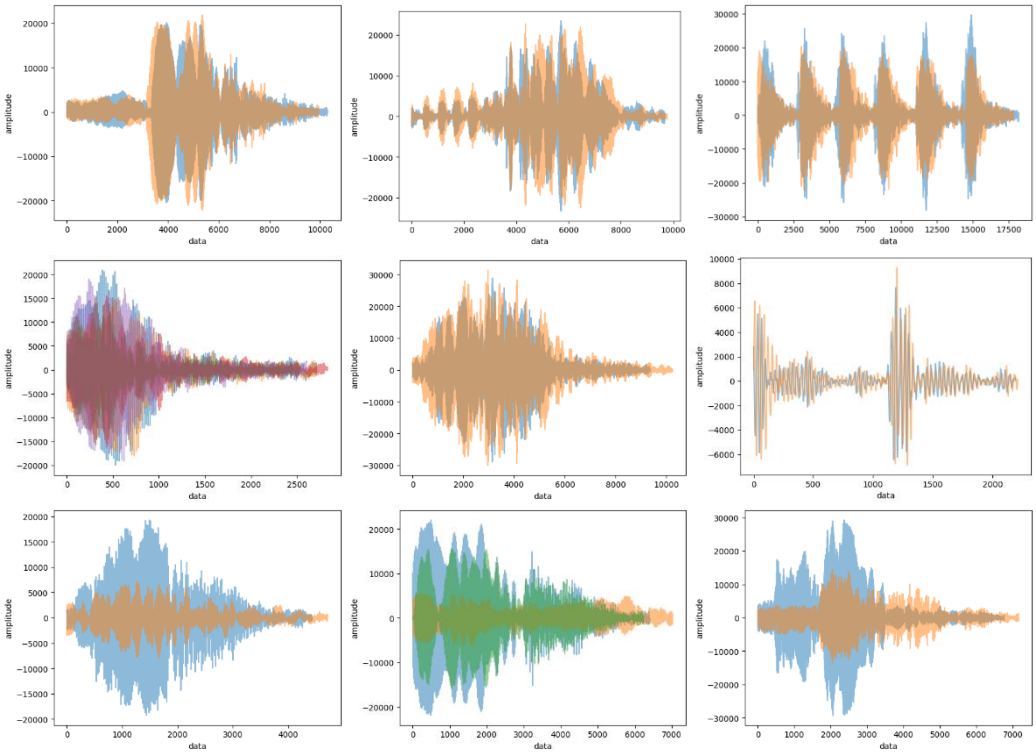


Fig. 8. Syllable Grouping Evaluation: Comparing 6 Well-Grouped Syllables with 3 Partially Grouped Syllables.

E. Syllabic Statistic

As we can see at Figure 9, a remarkable finding emerges when examining the occurrence frequency of syllables in an Indonesian passage and the birdsong of the Jalak Suren species. The statistical analysis of syllable frequencies reveals striking similarities between the two datasets. The blue line represents the

sorted syllable frequency distribution in the Indonesian passage, while the orange line represents the syllable frequency distribution in the Jalak Suren birdsong.

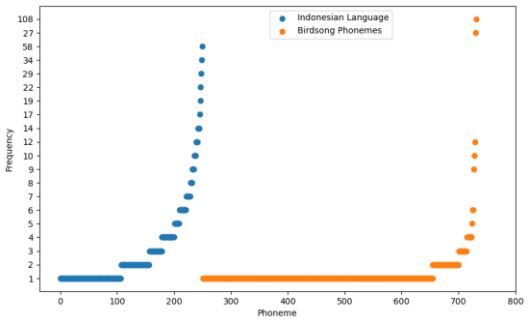


Fig. 9. Comparison of Syllable Occurrence Frequency: Indonesian Passage vs. Birdsong.

This observation provides compelling evidence that birdsong exhibits consistent syllabic patterns akin to human language usage. Further investigation is required to determine if these syllables form distinct words within the birdsong, which would not be unexpected based on the findings. This finding also suggests the possibility of detecting individual words uttered by birds and potentially mapping them to specific meanings in human language.

IV. CONCLUSION

Automatic syllable segmentation has been successfully applied to analyze the birdsong of the Jalak Suren species. The analysis reveals a significant number of repeated syllables characterized by remarkable similarity in terms of audio signal shape and duration of chirps. These findings indicate that Jalak Suren birds are not only sensitive to tone but also to syllable duration. Short syllables are likely to go unrecognized by these birds, reinforcing the notion that their vocalizations possess intentional meaning rather than being random. The analysis resulted in the identification of 481 unique syllables within the recorded birdsong sequence, nearly double the number found in the Indonesian language passage. Moreover, the statistical occurrence of syllables in birdsong displayed striking similarities to those observed in human language passages, suggesting the presence of organized language within the avian species.

Further research in this area should focus on several key aspects to deepen our understanding of birdsong and its linguistic properties. First, investigating the relationship between specific syllable patterns and their corresponding meanings would shed light on the underlying semantic structure of birdsong. This could involve further experimental studies that explore the responses of birds to different sequences of syllables and their associated behaviors or environmental contexts. Moreover, conducting comparative studies across different bird species and their respective vocalizations could uncover valuable insights into the universality or species-specific nature of certain phonetic or syllabic structures. Examining birds with diverse ecological and evolutionary backgrounds would help elucidate the evolutionary origins and adaptive functions of birdsong.

REFERENCES

[1] A. J. Doupe and P. K. Kuhl, “Birdsong and human speech: Common themes and mechanisms,” *Annual Review of Neuroscience*, vol. 22, pp. 567-631, 1999.

[2] J. Podos and P. S. Warren, “The Evolution of Geographic Variation in Birdsong,” *Advances in the Study of Behavior*, vol. 37, pp. 403-458, 2007. [Online]. Available: https://doi.org/10.1016/S0065-3454(07)37009-5.

[3] M. C. Baker and M. A. Cunningham, “The biology of bird-song dialects,” *Behavioral and Brain Sciences*, vol. 8, no. 1, pp. 85-100, 1985.

[4] T. N. Suzuki, D. Wheatcroft, and M. Griesser, “Wild Birds Use an Ordering Rule to Decode Novel Call Sequences,” *Current Biology*, vol. 27, no. 15, pp. 2331-2336.e3, 2017. [Online]. Available: https://doi.org/10.1016/j.cub.2017.06.031.

[5] S. Engesser, A. R. Ridley, and S. W. Townsend, “Meaningful call combinations and compositional processing in the southern pied babbler,” in *Proceedings of the National Academy of Sciences*, vol. 113, no. 21, pp. 5976-5981, 2016. [Online]. Available: https://doi.org/10.1073/pnas.1600970113.

[6] J. B. Nuckolls, "The Case for Sound Symbolism," *Annual Review of Anthropology*, vol. 28, no. 1, pp. 225-252, 1999. doi: 10.1146/annurev.anthro.28.1.225.
[7] S. F. Schmerling, Sound and Grammar: a Neo-Sapirian theory of language. Brill, 2018.
[8] K. de Jong, "Temporal constraints and characterising syllable structuring," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, J. Local, R. Ogden and R. Temple, Eds. Cambridge University Press, 2003, pp. 253-268. doi: 10.1017/CBO9780511486425.015.
[9] F. Levy, "Mirror neurons, birdsong, and human language: a hypothesis," *Front. Psychiatry*, vol. 2, pp. 78, Jan. 2012. doi: 10.3389/fpsyt.2011.00078.
[10] C. Bentz and D. Alikaniotis, "The word entropy of natural languages," *arXiv preprint arXiv:1606.06996*, 2016.
[11] T. Angguni, Y. A. Mulyani, and A. Mardiastuti, "Bird species contested at songbird competition in Jabodetabek Region, Indonesia," in *IOP Conference Series: Earth and Environmental Science*, vol. 762, no. 1, pp. 012014, 2021. [Online]. Available: https://dx.doi.org/10.1088/1755-1315/762/1/012014.
[12] "Google Trends for Jalak Suren." [Online]. https://trends.google.co.id/trends/explore?date=today%205-y&geo=ID&q=jalak,murai,lovebird,kenari&hl=id. [Accessed: 1-Jun-2023].
[13] B. Faghih and J. Timoney, "Smart-Median: A New Real-Time Algorithm for Smoothing Singing Pitch Contours," *Applied Sciences*, vol. 12, no. 14, pp. 7026, Jul. 2022, doi: 10.3390/app12147026.
[14] M. C. Yesilli, J. Chen, F. A. Khasawneh, and Y. Guo, "Automated surface texture analysis via Discrete Cosine Transform and Discrete Wavelet Transform," *Precision Engineering*, vol. 77, pp. 141-152, 2022. [Online]. Available: https://doi.org/10.1016/j.precisioneng.2022.05.006.
[15] B. G. Osgood, "Lectures on the Fourier transform and its applications," *American Mathematical Soc.*, 2019.
[16] L. F. Costa, "Further generalizations of the Jaccard index," *arXiv preprint arXiv:2110.09619*, 2021.