An analysis of the persistence of Zonotrichia capensis themes using dynamical systems and machine learning tools

Article *in* Chaos Solitons & Fractals · December 2022

DOI: 10.1016/j.chaos.2022.112803

Roberto Andrés Bistel Esquivel
Universidad de Buenos Aires

Gabriel Mindlin
Universidad de Buenos Aires

Chaos, Solitons and Fractals 165 (2022) 112803

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

An analysis of the persistence of *Zonotrichia capensis* themes using dynamical systems and machine learning tools

Roberto Bistel [a,b], Alejandro Martinez [b], Gabriel B. Mindlin [a,b,c,*]

[a] Departamento de Física, FCEyN, Universidad de Buenos Aires, Argentina
[b] IFIBA, CONICET, Buenos Aires, Argentina
[c] Departamento de Matemática Aplicada, Universidad Rey Juan Carlos, Madrid, Spain

A R T I C L E I N F O

Keywords:
Birdsong
Neural networks
Models of phonation

A B S T R A C T

In this work, we use tools from nonlinear dynamics to generate synthetic bird songs with frequency modulations compatible with sketches reported in a study of chingolo (*Zonotrichia capensis*) songs in 1966. Using machine learning tools, we conclude that some of the sketches correspond to themes that are still sang in the same region, five decades later.

1. Introduction

The chingolo (*Zonotrichia capensis*) is an oscine bird found in South America, some regions of Central America, and some Caribbean islands. The song of the chingolo includes a first part called theme. The theme consists of a few (one to five) notes with modulated frequency. Each chingolo has a particular theme, although there are chingolos that can sing two or three different themes. The second part of the chingolo song is the trill, which is a fast repetition of identical notes with decreasing frequency [1–3].

The trill of the chingolo varies between areas with different environments. The trill remains stable within an ecologically homogeneous area. F. Nottebohm made the first study of the incidence of song variation in some Argentine populations of chingolos. This study, conducted in 1966, made it possible to identify a set of different themes in various geographical locations. In [1], the author found examples of variations of chingolo themes in Parque Pereyra Iraola (Buenos Aires Province, Argentina). In this work, we analyze, five decades later, the persistence of *Zonotrichia capensis* themes reported in 1966 in Parque Pereyra. The aim is to test the hypothesis that chingolos in Parque Pereyra still sing some themes described by [1].

The themes cataloged by Nottebohm are a set of manual field annotations made from his auditory interpretation (without computing sonograms) of the songs. In [1], the author did not use standard equipment for recording the songs on tape or as spectrographs. Therefore, the only report of these themes are the field annotations described in [1]. In this paper, we propose to use tools from machine learning and dynamical systems to compare the themes reported in 1966 with data from field audio recordings made in the same area in 2020.

We trained and validated an artificial neural network which compares and learns the similarities between groups of songs of the same theme. The neural network takes the image of a spectrogram corresponding to a theme and generates a *3D* embedding. We used *PCA* to reduce the hyperspace to *2D* and create a graph of themes clusters. To train the neural network, we used only images of spectrograms corresponding to synthetic songs made through a computational model. The model for synthesizing songs brought into the present the themes from 1966 and performed multiple variations on the synthetic themes corresponding to the 2020 field recordings. We used these synthetic songs to train a neural network.

By processing our data with neural network, we identified that at least three of the *Zonotrichia capensis* themes reported in Parque Pereyra in 1966, persist in 2020. This is the first study to report the persistence of *Zonotrichia capensis* themes in the same area for more than three decades.

2. Identification of themes in the songs of *Zonotrichia capensis*

The common chingolo (*Zonotrichia capensis*) is an oscine bird that executes a stereotyped song, made up of notes (syllables) that follow a particular pattern. Two parts compose the song of chingolo: the introduction, and the trill. The introduction is a sequence of between two and five whistle notes (pure tones with slow frequency modulation). Depending on the number, configuration, and order of the notes in the

* Corresponding author at: Departamento de Física, FCEyN, Universidad de Buenos Aires, Argentina.
E-mail address: gabo@df.uba.ar (G.B. Mindlin).

https://doi.org/10.1016/j.chaos.2022.112803
Received 28 June 2022; Received in revised form 20 September 2022; Accepted 10 October 2022
Available online 20 October 2022

introduction, it is possible to distinguish different themes. The trill is a more or less rapid repetition of the same note within descending frequency modulation. Fig. 1 shows a spectrogram of a typical chingolo song, where we indicate the theme and the trill [1–3].

The features of the habitat determine important characteristics of the trill. Different trills vary in the number of notes, their duration, the inter-syllabic time separation, and the maxima and minima of the frequencies. A group of birds, in an ecologically homogeneous area, share these trill features, giving way to dialects [1–4].

In general, each chingolo has a specific theme, although there are chingolos that can sing two or three different themes. The features of a theme are distinctive to each bird, so they are used as identity signatures [5,6].

To identify the different themes, a trained researcher visually analyzes the spectrogram of the song and identifies some patterns of the sound [7]. In the chingolo, typical features that must be analyzed are the shape, the type of modulation, and the number of notes. These patterns can be classified broadly in four classes: linear, concave, convex, and concave-convex. The four types of frequency modulations are: ascending, descending, ascending-descending, and without modulation (constant frequency) [5].

F. Nottebohm conducted one of the firsts studies on the geographical variations of the chingolo's song in Argentina in 1966. The author did not use any audio recording system or sound-spectrographic analysis equipment, so there are no audio files of the reported songs or their spectrograms. The only report available of the songs is a set of annotations based on his estimation of the previously described sound features [1].

One of the areas that he studied was Parque Pereyra. The author cataloged twelve themes of chingolos that share the same dialect in the studied area [1]. Fig. 2 shows the representation of the themes reported, indicating each theme with an alphabetical letter.

In 2020, we recorded multiple chingolo songs along a 3 km internal road in Parque Pereyra (*GPS* coordinates: (−34.8616, −58.1163) to (−34.8659, −58.1382)). A preliminary analysis of the spectrograms allowed us to identify 13 different themes. In Fig. 3, we show their spectrograms. We computed each spectrogram after processing the audio recordings with a noise reduction filter and a band-pass filter with cutoff frequencies between 1.5 kHz to 8 kHz. We applied a Gaussian window (standard deviation of 128 points), processing segments of 1024 samples, with successive overlaps of 512 samples. To visualize these spectrograms, we considered a clipping of less than 1/600 of the maximum value of each spectrogram's signal [6].

A visual comparison between the themes of 1966, and the ones we measured in 2020, allowed us to identify that there are similar themes. However, similarity clustering using visual analysis is subjective, and depends on the researcher's experience identifying patterns as well as it is conditioned by his/her biases [7,8].

Machine learning techniques and neural networks are robust and

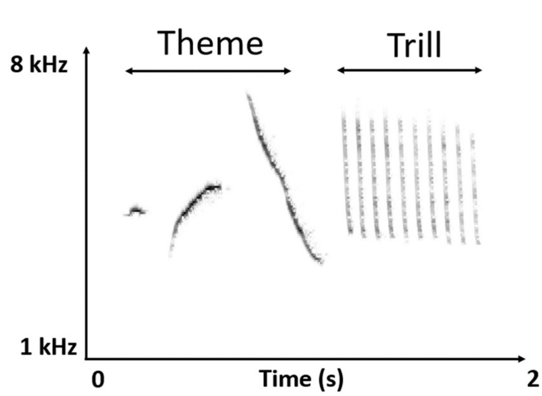

Fig. 1. Spectrogram of a typical chingolo song.

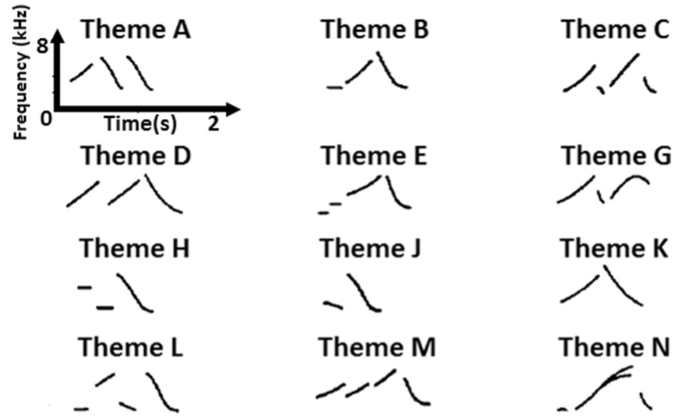

Fig. 2. Reported themes in Parque Pereyra in 1966 (adapted from [1]).

reproducible techniques used to analyze large volumes of data in an unsupervised or semi-supervised way. Convolutional neural networks are designed to efficiently extract the optimal features from image data that lead to efficient classification. Among the uses of neural networks are automatic classification and data clustering based on their similarities [9].

The training process of a neural network requires a previously classified representative data set. In our case, to use a neural network that organizes the chingolo themes according to their similarities, we need many songs corresponding to each theme. Typical data augmentation operations used in machine learning tasks do not apply to this research, as they generate songs variations that are inconsistent with biological variability [5,6]. Furthermore, there are no spectrograms from the songs described in 1966.

We used a computational model for synthetic song generation in order to produce multiple variations of chingolo songs. We obtained these variations by modifying the model parameters within a biologically acceptable range [5,6]. In this way, we created a set of synthetic data corresponding to the spectrograms of each theme. We used only images of synthetic spectrograms to train a neural network capable of clustering the themes based on similarity estimates.

3. The computational model for synthesizing songs

The computational model of synthetic song generation describes the physics of birdsong production. The sound-producing organ in birds is the syrinx, which is a structure that supports two pairs of lips, at the junction between the bronchi and the trachea. When a strong enough airflow passes through the lips, they vibrate, producing oscillations that modulate the airflow and generate the emitted sound [5,10,11].

Singing occurs through the control of air sac pressure and the configuration of the syrinx. Pressure control allows the airflow through the lips to be varied, while the configuration of the syrinx affects the fundamental frequency of the oscillations. The increased air sac pressure provokes the threshold of the oscillatory movement of the lips [5,10,11].

The model assumes the lips are stationary when the bird is silent. The airflow modulated through an increased pressure causes it to remain in the phonation region. A reduction of the pressure causes the sound to stop. The disappearance of the oscillations occurs in an inverse Hopf bifurcation. In the proximity of the bifurcation region, the model transforms into a set of equations describing the labial dynamics. Eq. (1) shows the system of equations that describes the labial dynamics for *Zonotrichia capensis* [5,11].

$$\begin{cases} \dfrac{dx}{dt} = y \\ \dfrac{dy}{dt} = k\gamma^2 x - \gamma x^2 y + \beta\gamma y \end{cases} \tag{1}$$

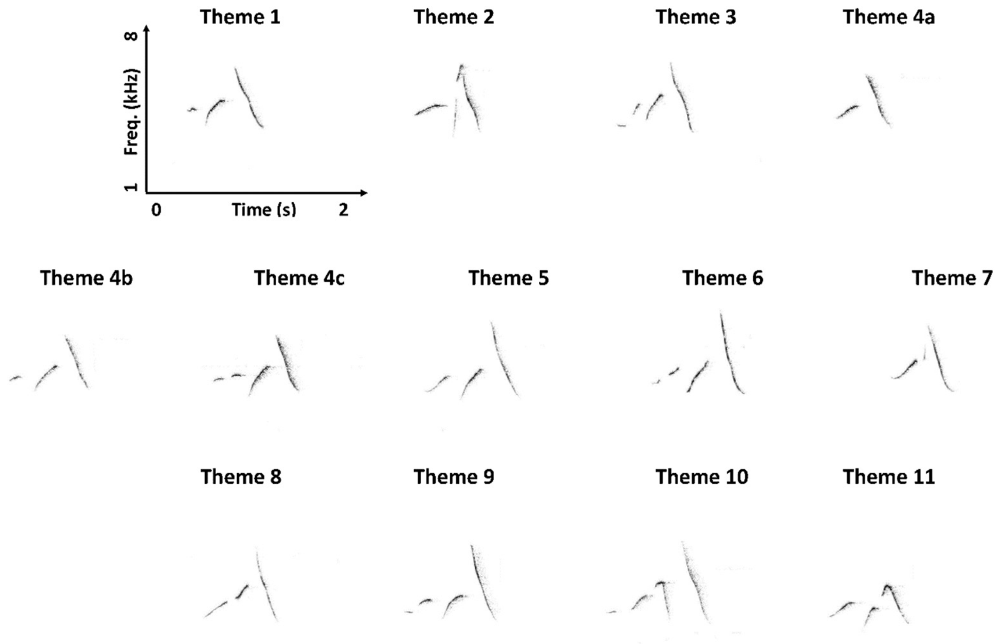

Fig. 3. Spectrogram of chingolo themes in 2020.

In Eq. (1), x represents the midpoint labial position; k is the restitution constant, which is proportional to the labia tension; β is a negative dissipation, proportional to air sac pressure; while γ is the timescale of the system. Eq. (2) models the pressure at the tracheal input p_i needed to generate sound [5].

$$p_i(t) = A\,\frac{dx(t)}{dt} + p_{back}\left(t - \frac{L}{c}\right)$$

$$p_{back}(t) = -r\,p_i\left(t - \frac{L}{C}\right) \tag{2}$$

In Eq. (2), A is a coefficient that depends on the airflow, L is the length of the trachea, c is the speed of sound in the medium, and r is the reflection coefficient. The pressure at the output of the trachea is $p_o = (1-r)p_i\left(t - \frac{L}{C}\right)$, which forces the neck of a Helmholtz resonator which represent the oropharyngeal-esophageal cavity (OEC). The set of Eq. (3) models the operation of the OEC as a signal filter [12].

$$\frac{di_1}{dt} = i_2$$

$$\frac{di_2}{dt} = -\frac{i_1}{cL_1} - \left(\frac{r_d}{L_2} + \frac{r_d}{L_1}\right)i_2 + \left(\frac{1}{cL_1} + \frac{r_d r_2}{L_1 L_2}\right)i_3 + \frac{\frac{dp_o}{dt}}{L_1} + \frac{r_d r_2}{L_1 L_2}p_o \tag{3}$$

$$\frac{di_3}{dt} = -\left(\frac{L_1}{L_2}\right)i_2 - \left(\frac{r_d}{L_2}\right)i_3 + \left(\frac{1}{L_2}\right)p_o$$

An equivalent electrical circuit represents the dynamic of the Helmholtz resonator with an opening. The sound emitted is proportional to the value of i_3, ruled by a set of equations derived in [12]. The parameter's values corresponding to eq. 1 are $\gamma = 24000$; while the standard value of $\beta = -0.15$. The β value follows a normal distribution $(0, 0.05)$ and is varied in every integration of the model. In Eqs. (2) and (3), the parameters used to generate synthetic songs in *Zonotrichia capensis* are $L = 0.025$; $r = -0.65$; $L_1 = \frac{1}{20}$; $L_2 = \frac{1}{10^4}$; $r_2 = 0.5 \times 10^7$; $r_d = 24 \times 10^3$; and $c = \frac{5}{350 \times 10^8}$ [5].

The complexity of the acoustic modulations of singing arises from simple modifications of a very generic gesture in lung pressure and vocal fold tension [13]. The acoustic modulations in *Zonotrichia capensis* correspond to three elementary frequency modulation patterns: sinusoidal, linear, and exponential downsweep. Table 1 shows the parameters for each modulation pattern.

The model takes as input the modulation pattern of each note and the necessary parameters for its reproduction (Table 1). Then, the model generates a list of fundamental frequencies for each note. The value of the parameter k, that allows obtaining the fundamental frequency w, is $k = 6.5 \times 10^{-8}w^2 + 4.2 \times 10^{-5}w + 2.6 \times 10^{-2}$. The relationship between k and w was obtained through a series of numerical simulations in the parameter space of the model, varying the values of k and computing for each simulation the fundamental frequency w of the synthesized song. In this way, we propose a polynomial relationship between k and w, to then use the list of pairs (k, w) to compute the coefficients of the polynomial through a numerical regression [11]. Then, the model uses the list of fundamental frequencies transformed into the necessary parameters to synthesize a song. The final song retains the spectral content of the sound emitted by the bird.

To reproduce the basic gestures, we integrated the dynamic model many times, varying the values of the parameters as shown in Table 1. We obtained the range of variation of the parameters from the statistical analysis of song samples of each theme identified in the field recordings of 2020. We analyzed the spectrograms of ten songs per theme, except for *Theme 4c*, for which there are only two songs recorded.

The parameters that characterize each song (the initial and final values of the fundamental frequency of each note, the duration of each note, and the time difference between the notes) have variation in a range of less than 7 % between the different songs. To obtain the

Table 1
Basic frequency modulation patterns for the song of the *Zonotrichia capensis.*

Modulation pattern	Frequency	Parameters
Sinusoidal	$w(t) = w_f + (w_i - w_f)\left(\frac{t - t_i}{t_f - t_i}\right)$	w_f, w_i, t_f, t_i
Linear	$w(t) = w_{av} + A\sin(\alpha_i + (\alpha_f - \alpha_i))\left(\frac{t - t_i}{t_f - t_i}\right)$	$w_{av}, A, \alpha_f, \alpha_i, t_f, t_i$
Exponential	$w(t) = w_f + (w_i - w_f)e^{-\frac{3(t - t_i)}{(t_f - t_i)}}$	w_f, w_i, t_f, t_i

variations of the values in the parameters, we calculated a Gaussian distribution with mean and standard deviation from the song examples for each of the themes.

To generate the themes of 1966, we analyzed the patterns of all the themes (i.e. those from 1966 as well as the ones recorded in 2020). Among all the themes, we identified the presence of notes that follow patterns of the linear type of constant frequency, ascending linear, rising exponential and falling exponential. We considered the ascending-descending type patterns as a combination of an ascending linear gesture followed by a falling exponential gesture.

To obtain the range of variations of the parameters used to synthesize the songs of 1966, we grouped the similar patterns of all song recorded in 2020. In this way, we have a pattern conformed by songs of different themes. This result in a range of variation for each pattern of 12 %.

The synthetic spectrograms presented random differences in their parameters, consistent with biological variability among the songs of a theme. We generated 100 synthetic spectrograms for each theme as surrogate data. From the themes of the year 2020, we randomly took 60 spectrograms for training the neural network, and 20 spectrograms for the validation. We separated this sample of 80 spectrograms from the remaining 20 spectrograms, to later use these last data in the network testing. Fig. 4 shows examples of the spectrograms generated by the computational model for some themes from 1966 and 2020.

4. Neural network for themes identification

The neural network used is a siamese type, where the cost function is the contrastive loss. We built a siamese network because it allows us to employ a small data set and determine the similarities between different classes (themes). We took images of synthetic spectrograms corresponding to the year 2020 themes to train and validate the neural network. The network clustered each spectrogram based on similarities, maximizing the distance between groups of spectrograms that correspond to different themes. In the tests, we analyzed the similarities of the themes of the year 2020 with those of 1966. We say that a chingolo theme persists if the distances between clusters are so small that they overlap in space.

A siamese network is built from two identical branches that share the same architecture and weights. Each branch is a convolutional neural network that uses as input the images of the synthetic spectrograms. Convolutional neural networks are designed to extract the optimal features from image data that lead to efficient classification. Neural networks include a set of fully connected convolutional layers and *ReLU* activation functions. The objective of the contrastive loss function is to determine the Euclidean distance between the features obtained by introducing the two images into the model [14]. In this way, we built clusters of similar themes and maximized the distance between groups of different themes.

The use of synthetic spectrograms as input of the neural network allows us to use the trained network with images of spectrograms of field recordings. To obtain the spectrogram of field recording, we use different strategies to extract/emphasize the spectral pattern of the sound. We can expose these patterns, adjusting the time-frequency resolution used to compute the spectrogram.

We built the convolutional networks with four *2D* layers alternating with four *Max-Pooling* layers. The network features a final pair of dense layers. The *2D* convolutional layers have output filter sizes of 64, 128, 256, and 512, obtained from their inputs after performing a convolution with 15×15, 7×7, 4×4, and 4×4 size windows, respectively. All *Max-Pooling* layers perform a dimensionality reduction by a factor of 2, which makes the images smaller. This allowed to reduce the computational cost, minimize the possibility of *overfitting* and increase the abstraction of the input data. The final two dense layers consisted of 1024 and 3 units, respectively. The final layer had 3 units because we want to represent the data in a *3D* space.

In the early stages of the work, we train the network using an output layer of two neurons, thus obtaining a direct representation in a *2D* plane. The trained models with two output neurons presented overlapping songs of different themes. In these models, the distances between clusters are small.

To increase the resolution and the distance between clusters, we increase the output dimension and train the networks using an output layer of three neurons. We use the *PCA* method for the *2D* visual representation of the output data.

To avoid overfitting, we set the regularization parameter $l2 = 0.0002$, and we randomly dropped some weights of the network connections (setting their values to zero). We set the dropout value to 0.2 and the learning rate to 0.001. We converted our images to grayscale format, with a size of 300×200 pixels. We set the batch size as 128 units, and the training took place in 450 epochs. We trained the network with the *Keras* library and the *ImageDataGenerator* class, in which each image was a tensor. We normalized the values in each image to 255. The tensor output represents the distance between the input images to the network. A smaller distance means more similarities between the images.

The contrastive loss function uses *Keras/TensorFlow*. The input arguments are a pair of images with their labels and the margin value for the loss function. The labels allow setting a value equal to ˝1˝ when the two images are similar, and ˝0˝ when they are of different classes. Eq. (4) shows that the function returns the loss value for a pair of input images [14,15].

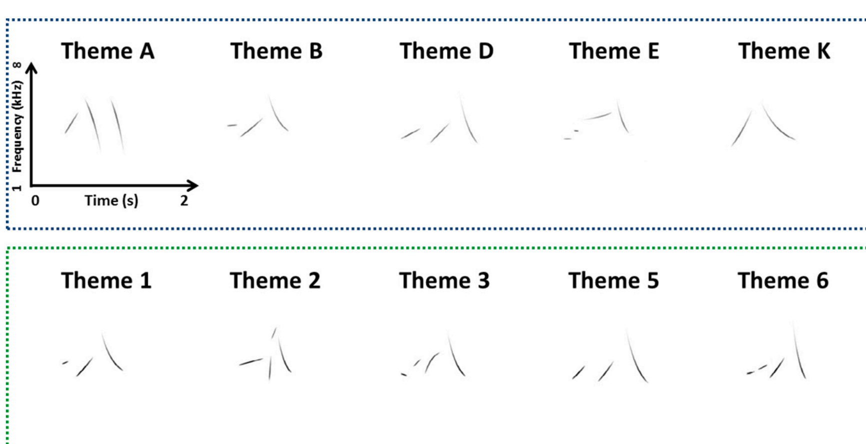

Fig. 4. Spectrograms generated by the computational model. The blue box indicates 1966 themes and the green box 2020 themes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$L_n = \frac{1}{2}\left(y_n {d_n}^2 + (1 - y_n) max(margin - d_n, 0)^2\right) \tag{4}$$

In Eq. (4), the variable *y* represents the assignment of a value ˝1˝ or ˝0˝ in correspondence with the similarities of the pairs of images; d_n is the Euclidean distance between the vector of features learned by the network. For a couple of images f_0 and f_1, $d_n = \|f_{0_n} - f_{1_n}\|_2$. The term *margin* reinforces the constraint: if two images in a pair are different, their distances must be at least the value set for the *margin*, or a loss will occur. The *margin* value in our function is equal to 2.

5. Results of the analysis of the themes

We used the network to analyze the similarities between the themes of 2020 with those described for 1966 in Parque Pereyra. We trained the neural networks ten times, so we ended up with ten models. Each model uses a random set of synthetic spectrogram images. To test the neural network, we used the set of 20 images of synthetic spectrograms that we separated from the training and validation set.

We performed a test of similarities for each of the ten models of the neural network. Then, we projected the location of the songs on a *2D* plane. Since the output of the network was a *3D* vector, we used *PCA* to reduce the dimensionality to *2D*. The maximum value of sample variance provided by the third component was 1.2 %, so the *2D* representation is reliable. Fig. 5 shows the *2D* visualization of the results of the similarities analysis of the chingolo themes for one of the trained models.

In Fig. 5, we indicate with black arrows three areas where there is an overlap of themes. *Arrow* 1 points to a cluster of *Theme* 1 (2020) and *Theme* 4*b* (2020) with *Theme B* (1966). The overlap between these themes allows us to conclude that *Theme B*, *Theme* 1, and *Theme* 4*b* are the same. Therefore, *Theme B* persists. We need to consider *Theme* 1 and *Theme* 4*b* as the same theme, since their statistical parameters and modulation patterns are similar. The main difference between these themes is in the duration and the value of the frequency of the first note, varying 5 % between the two themes. During the visual identification of the themes of 2020, we consider *Theme* 4*b* as different, since it is sung by a chingolo capable of executing three different themes [6].

In Fig. 5, we also indicate with *Arrow* 2 the cluster of *Theme D* with *Theme* 5; and with *Arrow* 3 the cluster of *Theme K* with *Theme* 4*a*. *Theme D* and *Theme* 5 are themes that share three notes, where the first two are ascending linear-type patterns, and the third note is an exponential downsweep. *Theme K* and *Theme* 4*a* have two notes, the first is an ascending exponential pattern, and the second is the downsweep.

To quantify similarities, we computed the distance separating the themes of 1966 from those of 2020. We compute the distance in *2D* after applying the *PCA* method. A chingolo theme persists if the distance separating the clusters is small, and therefore an overlap occurs. In Fig. 6, we show the distance between *Theme B* and the identified themes in 2020. The data shown in Fig. 6, was computed with the data used to generate Fig. 5.

The closest distance to *Theme B* corresponds to the songs of *Theme* 1 and *Theme* 4*b*. Moreover, Fig. 6 shows the overlap between the songs in *Theme* 1 and *Theme* 4*b*. In all the trained models of the neural network, the songs in *Theme B* are always at a smaller distance from the songs of *Theme* 1 and *Theme* 4*b* in respect to the other themes.

In Fig. 7, we show that for the ten trained models of the network, the distance between *Theme B* to *Theme* 1 and *Theme* 4*b* remains constant. The trained models allow us to compute the mean distance and the standard deviation between a theme to the rest. In Table 2, we show the mean distances and standard deviation values for *Theme B* to *Theme* 1 and *Theme* 4*b*; *Theme D* to *Theme* 5; *Theme K* to *Theme* 4*a*.

The distance analyses for the multiple trained models allows us to conclude that at least three themes of *Zonotrichia capensis* reported by [1] in 1966 persist in Parque Pereyra in 2020. In this work, we report for the first time the persistence of chingolo themes for more than three decades in a micro geographical area.

We explain the persistence of these themes from the point of view of the learning process of the juveniles, since it is probable that they take as a reference the themes more sung in an area. The themes mostly sung in Parque Pereyra in 1966 are *Theme B* and *Theme D*, with 74 % of the chingolos singing these themes, while 7 % sang *Theme K* [1]. *Theme B* is also the most sung in the northeast of Buenos Aires, according to a study made in 1990 [16]. Therefore, the themes that have persisted are those sung by the majority of chingolos.

6. Conclusions

In this work, we analyze, five decades later, the persistence of *Zonotrichia capensis* themes identified in Parque Pereyra in 1966. To analyze the persistence of the themes, we implemented an automatic processing method by means of a siamese neural network with a contrastive loss function. In order to train, validate and test the neural network, we used images of spectrograms computed from synthetic songs generated by a computational model. Using these techniques, we report for the first time the persistence of chingolo themes for more than three decades in a micro geographical area. According to [1], the songs

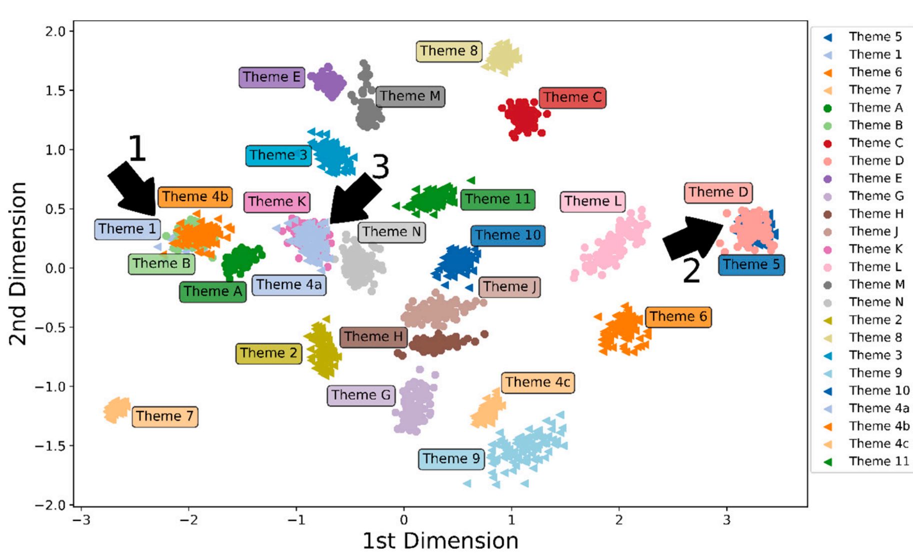

Fig. 5. Similarities analysis for one of the trained models.

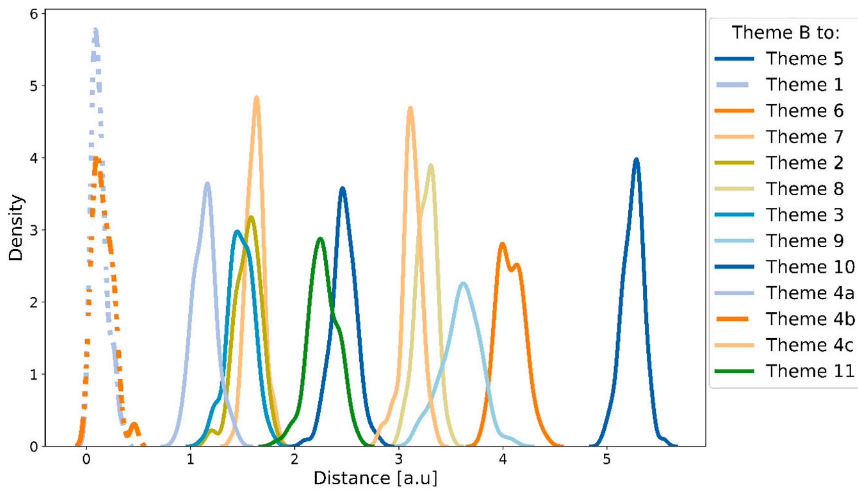

Fig. 6. Distances between *Theme B* songs to 2020 theme songs.

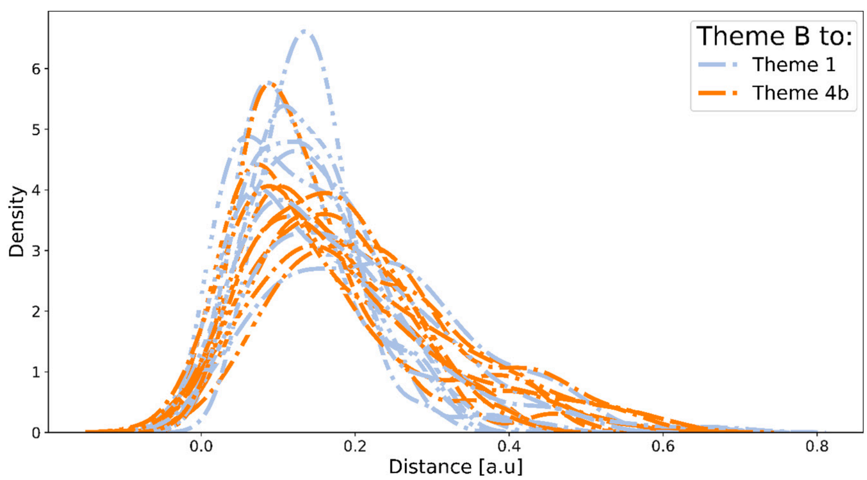

Fig. 7. Distances between songs of *Theme B* to songs of *Theme* 1 and *Theme* 4*b* for all models trained.

Table 2
Values of mean distance and standard deviation for the themes that persist in 2020.

	Mean [a.u.]	σ [a.u.]
Theme B to *Theme* 1	0.155	0.097
Theme B to *Theme* 4b	0.179	0.115
Theme D to *Theme* 5	0.143	0.087
Theme K to *Theme* 4a	0.112	0.069

that are preserved are those that were sung by most of the chingolos in Parque Pereyra in 1966.

A juvenile learns a theme through the imitation of the song of the adults of their population. The annual disappearance rate for adult chingolos ranges from 30 to 77.6 %. Then, a theme has a greater probability of persisting if a higher number of juveniles learn a theme [17].

In the last years, machine learning tools have allowed us to carry out data driven projects that were impossible just a few years ago. In this work, we used computational nonlinear models of song production to bring back to life songs compatible with sketches made five decades ago. We also used machine learning tools to compare them with present recordings. The machine learning tools used in this work, which are siamese neural networks, allow a comparison between the songs without the need to design the features that are pertinent for the comparison. This is an example of how nonlinear dynamics and machine learning can provide us with tools that will enrich our studies of animal vocal communication.

CRediT authorship contribution statement

Roberto Bistel collected and analyzed the data, and participated in the writing the manuscript, Alejandro Martinez participated in the data collection, Gabriel Mindlin participated in the collection and analysis of the data, in the conception of the work, and participated in the writing of the manuscript.

Declaration of competing interest

The authors declare no conflict of interests.

Data availability

Data will be made available on request.

Acknowledgements

GM acknowledges URJC for the hospitality and support during his sabbatical stays. This work was partially supported by UBA and ANCyT.

References

[1] Nottebohm F. The song of the chingolo, Zonotrichia capensis, in Argentina: description and evaluation of a system of dialects. The Condor 1969;71(3): 299–315.

[2] King JR. Variation in the song of the rufous-collared sparrow, Zonotrichia capensis, in northwestern Argentina. Z. Tierpsychol. 1972;30(4):344–73.

[3] Kopuchian C, Lijtmaer DA, Tubaro PL, Handford P. Temporal stability and change in a microgeographical pattern of song variation in the rufous-collared sparrow. Anim. Behav. 2004;68(3):551–9.

[4] Mundinger P. Microgeographic and macrogeographic variation in acquired vocalizations in birdsKroodsma DE, Miller EH, editors. ‘Acoustic Communication in Birds. Vol. 2: Song Learning and its Consequences’ 1982:147–208.

[5] Tubaro PL, Mindlin GB. A dynamical system as the source of augmentation in a deep learning problem. Chaos, Solitons & Fractals: X 2019;2:100012. https://doi.org/10.1016/j.csfx.2019.100012.

[6] Bistel RA, Martinez A, Mindlin GB. Neural networks that locate and identify birds through their songs. The European Physical Journal Special Topics 2021;1–10. https://doi.org/10.1140/epjs/s11734-021-00405-5.

[7] Giret N, Roy P, Albert A, Pachet F, Kreutzer M, Bovet D. Finding good acoustic features for parrot vocalizations: the feature generation approach. J. Acoust. Soc. Am. 2011;129(2):1089–99. https://doi.org/10.1121/1.3531953.

[8] Jones AE, ten Cate C, Bijleveld CCJH. The interobserver reliability of scoring sonograms by eye: a study on methods, illustrated on zebra finch songs. Anim. Behav. 2001;62(4):791–801. https://doi.org/10.1006/anbe.2001.1810.

[9] Chollet F. Deep learning with python. Simon and Schuster; 2021.

[10] Goller amp, Suthers.. Role of syringeal muscles in controlling the phonology of bird song. J.Neurophysiol. 1996;76(1):287–300.

[11] Laje R, Gardner TJ, Mindlin GB. Neuromuscular control of vocalizations in birdsong: a model. Physical Review E. 2002;65(5):051921. https://doi.org/10.1103/PhysRevE.65.051921.

[12] Perl YS, Arneodo EM, Amador A, Goller F, Mindlin GB. Reconstruction of physiological instructions from zebra finch song. Phys. Rev. E. 2011;84(5):051909. https://doi.org/10.1103/PhysRevE.84.051909.

[13] Gardner T, Cecchi G, Magnasco M, Laje R, Mindlin GB. Simple motor gestures for birdsongs. Phys. Rev. Lett. 2001;87(20):208101. https://doi.org/10.1103/PhysRevLett.87.208101.

[14] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05) 2005;1:539–46. https://doi.org/10.1109/CVPR.2005.202.

[15] Rosebrock A. January 18). Contrastive Loss for Siamese Networks with Keras and TensorFlow. 2021:1–35. https://pyimagesearch.com/2021/01/18/contrastive-loss-for-siamese-networks-with-keras-and-tensorflow/.

[16] Tubaro PL. Aspectos causales y funcionales de los patrones de variación del canto del chingolo (Zonotrichia capensis). 1990.

[17] García NC, Arrieta RS, Kopuchian C, Tubaro PL. Stability and change through time in the dialects of a neotropical songbird, the rufous-collared sparrow. Emu-Austral Ornithology 2015;115(4):309–16. https://doi.org/10.1071/MU14099.