Ciência e Natura, Santa Maria, v. 36 Ed. Especial II, 2014, p. 646-654
Revista do Centro de Ciências Naturais e Exatas - UFSM
ISSN impressa: 0100-8307 ISSN on-line: 2179-460X

CIÊNCIA e NATURA 35 anos

Automatic classification of *Turdus rufiventris* song notes by spectrographic image template matching

Classificação automática das notas do canto do *Turdus rufiventris* por comparação de imagens espectrográficas

Nilson Evilásio Souza Filho[1], Beatriz Cerimeli Oliveira[2], Maria Luisa da Silva[3], Jacques Vielliard[4]

[1]Doutor, Eng Acústica, Universidade Federal de Santa Maria,RS, Brasil.
[2]Eng Ambiental - UEM
[3]Lab. Bioacústica - UFPA
[4]IB-Dep Zoologia-Unicamp; (IN MEMORIAN)

Abstract

This paper presents a method for automatic classification of birdsong notes. The elaborate method performs correlation calculations applied to spectrographic images to determine the similarity between the notes of a vocal repertoire, so the method was called SITM (Spectrographic Image Template Matching). The notes N_j of each phrase P_i which makes up the song of an individual of a specie, are generally classified with letters of the alphabet according to the sequence emitted that is viewed in a spectrogram. By setting a minimum degree of similarity between the notes of the repertoire of each individual, it is possible to automate this type of visual classification. Performance evaluation of automatic classification was performed with the vocal repertoire of three individuals of T. ruviventris recorded in to the wild. With 96% confidence, according to statistical inference, the rate of correct classification via SITM was 77% to 97% of the notes in recording FNJV5915, 64% to 82% of the notes in recording FNJV5932 and 74% to 97% of notes in recording FNJV5955.

Keywords: *Biacoustic, Rufous-bellied Thrush, spectrogram, template-matching, classification.*

Resumo

Este trabalho apresenta um método para a classificação automática de notas de canto de pássaros. O método elaborado executa cálculos de correlação aplicados a imagens espectrográficas para determinar a similaridade entre as notas de um repertório vocal, por isso o método foi chamado de SITM (Spectrographic Image Template Matching). As notas N_j de cada frase P_i que compõe o canto de um indivíduo de uma espécie, geralmente são classificadas com letras do alfabeto de acordo com a sequência emitida que é visualizada num espectrograma. Ao definir um o grau mínimo de similaridade entre as notas do repertório de cada indivíduo, é possível automatizar esse tipo de classificação visual. A avaliação do desempenho da classificação automática foi realizada com o repertório vocal de três indivíduos da espécie T. ruviventris gravados na natureza. Com confiança de 96%, segundo a inferência estatística, o índice de acerto da classificação via SITM foi de 77% a 97% das notas da gravação FNJV5915, 64% a 82% das notas da gravação FNJV5932 e 74% a 97% das notas da gravação FNJV5955.

Palavras-chave: *Biacústica, Sabiá laranjeira, espectrograma, template-matching, classificação.*

nilson.evilasio@eac.ufsm.br

Recebido: 18/02/2014 Revisado: 31/03/2014 Aceito: 23/04/2014

1 Introduction

The interest of man in the vocal sounds and screams of animals always existed, a clear example of this interest is the importance of onomatopoeia in indigenous languages (Berlin and O'Neill, 1981). The bioacoustics studies the various aspects of sound communication and, like any area of knowledge, depends on two requirements to develop: a consistent conceptual basis and a tool technically appropriate (Vielliard, 2000).

Thus, bioacoustics take benefit from the technology of recording and analysis of sounds which allows the communication signal must be easily maintained and defined in terms of physical parameters. The conservation of communication signals recorded in nature, provides an obvious help to ornithology of field, since it allows to understand various aspects of sound communication in birds (Vielliard, 1987).

Therefore, as modern scientific field, the bioacoustics has been developed in a relatively recent way within ornithology, which is one of the traditional areas of knowledge, with fundamental contributions in establishing the basic concepts of evolution, biogeography, taxonomy, ethology and biology of conservation (Vielliard, 1987).

The comparison and classification of animal sounds is a common task in bioacoustics research. Scientists perform a detailed comparison of sonograms and search relationships between the structure of sound in time-frequency plane and a diversity of variables of extrinsic context. The interest may be in the identification of relations between the habitat, the social context in which the sounds are used, or the question of whether particular songs that are recorded from a single individual or social group. If geographic populations are structurally similar to other different individuals, groups or populations.

A visual comparison and classification of sonograms by a trained human observer is a widely used method for classifying spectrograms in groups (Pisoni et al., 1983). Although the visual criteria is rarely specified, and the fact that multiple observers converge to similar schemes of classification suggests that structural majority of the birdsong is identified.

A crucial point in this type of study is the method used for the comparison and structural categorization of a set of sounds or signs. The criteria for selecting measurements frequently include ease and speed of the measure, significance in priority studies, or probable relevance to specific conjectures. When considering all these criteria, Clark et al. (Clark et al., 1987) developed a method of comparison of sounds which was called Spectrographic Cross-correlation (SPCC), wherein two sounds in a spectrogram are correlated and the peak value of correlation is regarded as a measure of sound similarity (Chen and Maher, 2006).

We propose an alternative method based on correlation calculations applied to images, a technique knows as Template Matching. As it is spectrographic image the method was called Spectrographic Image Template Matching (SITM). The SITM stores correlation data between notes and used it for automate the process of classification among them and determine the measurable amount of information (in *bits*/*symbol*) according to the sequence of notes issued during a sound communication process. The intent of these types of analysis (SPCC and SITM) is including all structural features discernible from two spectrograms rather than comparing a predetermined limited and possibly incomplete set of measurements.

The automated classification of notes is extremely useful because of the precision and the large number of data processed almost simultaneously generating results that a human observer could never achieve so quickly. An appropriate automatic classification can help in the study of long and varied birdsongs, as the case of Nightingale (Muscicapidae: Luscinia flaba megarhynchos) that can emit 200 kinds of notes, or in the analysis of dialects of Rufous-collared Sparrow (Emberizidae: Zonotrichia capensis) (Avelino and Vielliard, 2004), the fast songs of Blue-black Grassquit (Emberizidae: Volatinia jacarina) (Fandiño-Mariño and Vielliard, 2004), or to establish relationships with brain nuclei responsible for gene behavior expression as performed with hummingbirds (Jarvis et al., 2000).

2 Bioacoustic basic concepts

2.1 Field Research

In field research, is necessary the location of the species in their habitat. This localization can be done visually or auditorily by using binoculars lens, and with the use of GPS and maps for demarcating the spatial distribution of the species. After the localization, is done the recording and playback. A communication signal is recorded, so playback is performed to recognize other individuals of the same species and respond to simulated signal and thus new songs can be recorded.

2.1.1 *Rufous-bellied Thrush*

The animal of interest in this study is the Rufous-bellied Thrush (*Turdidae: Turdus rufiventris*), a species of bird very common and conspicuous in Neotropical fauna. He lives in the forest (mesophytic forest, secondary forest), edges of rainforest, parks, gardens and even in the city center when there is some gardening (Da Silva et al., 2000). The *T. rufiventris* is abundant and widespread, which allows the study of many individuals from different localities.

Rufous-bellied Thrush occurs in eastern and Brazil central, of Maranhão until Paraíba, Rio Grande do Sul and Mato Grosso, Uruguay and Paraguay, and neighboring regions of Bolivia and Argentina (Da Silva et al., 2000).The physical appearance, orange belly, beak and bright yellow in orbicular region contrast with the clear throat streaked with black, is not as extraordinary as his singing, melodic, varied and mainly responsible for its popularity in Brazil (Da Silva et al., 2000). The song of the Rufous-bellied Thrush is considered one of the most complexes, consisting of whistles and trills of mean height (frequency range from 1 to 4 kHz). The phrases are articulated in long sequences of regularly spaced notes and issued successively. Some individuals tend to issue short notes largely modulated by the end of the phrase (Da Silva et al., 2000).

2.2 Animal Sound Communication

As in any communication system, the acoustic signals of animal communication need to keep the information they carry throughout the three stages of communication: emission, transmission and reception. The signal should therefore arrive at the receiver so that its function is captured and identified. The emission and reception of a signal of communication are conducted through specialized anatomical and physiological capacities that are appropriate to the modality of the signal and the environment they live in the animal in question (Da Silva et al., 2000).

The communication between birds is established by song, which may be a successive repetition of calls or complex patterns of continuous sound units, the notes, which form long and repeated phrases or themes. The birdsong is unique for each species and some of them have regional dialects. It is temporarily limited by seasons and generally executed only by males. It is mainly used to announce and define the territory and attract female. The basis of animal communication, therefore, corresponds to the biological species concept.

The divergence in the establishment of criteria for the definition of communication feeds the discussion about how to assess a given situation and verify that the individuals involved communicate or not. Practical examples show how difficult it is to arrive at a definition of communication is unambiguous and which can be easily applied to all cases (Da Silva et al., 2000). Therefore, in this study, the word song (or birdsong) is used to designate the sign of vocal communication which has the primordial biological function of specific recognition (Vielliard, 1987).

The communication is often reaching predetermined outcomes. But for that, is need to set goals and measurable terms in which the result is wanted to achieve with communication.

It should be established, in behavioral terms (desired responses) which will indicate whether there has been a change in behavior desired and, therefore, the communication was efficient. When this does not occur, it is said that communication failed. One of the approaches used is the measure of information quantity in communication through the Shannon entropy (Da Silva et al., 2000).

But to do this type of analysis, it is first necessary to classify all the notes of the species under study. Consistent classification of spectrograms should involve more than one observer. To describe the intensity of agreement between two or more judges, or between two classification methods, has been used Kappa measure, which is based on number of concordant responses, i.e., the number of cases the result is the same between judges (Valentim et al., 2010), (Gama et al., 2011). The essence of SITM is a value called the degree of similarity, which is the calculation of correlation associated with the index of agreement between visual and auditory classifications.

2.3 Time-frequency analysis

Acoustic signals are generally not stationary. For the analysis of non-stationary signals, commonly used is a method which describes the signal in both time and frequency. As the name implies, the time-frequency analysis associates a temporal signal (one-dimensional function of time) with an image (a two-dimensional function of time and frequency) which shows spectral components of the signal as a function of time. Conceptually, it is possible think in this mapping as a time-varying spectral representation of the signal. This representation is analogous to a musical score, with the two main axes represented by time and frequency.

The value of time-frequency representation of signal provides an indication of the specific times at which certain spectral components of the signal are observed. One of the tools that are used in processing non-stationary signals, with a multitude of applications in audio, is the Short Time Fourier Transform (STFT).

The STFT of a signal $x(t)$ denoted by $X(\tau,\omega)$ is defined as the Fourier transform of the windowed signal, i.e.:

$$X(\tau,\omega) = \int_{-\infty}^{\infty} x(t) * w(t-\tau) e^{-j(2\pi f)t} dt \qquad (1)$$

Where $x(t)$ is a sign displayed by a temporal window $w(t)$ of limited extent and the parameter τ is the central position of the window. Many different types of windows are used in practice. Typically, they are symmetric, unimodal and, in general, have complex value (Cohen, 1995).

When considering a pair of purely sinusoidal signals whose frequencies are spaced with an angular separation $\Delta\omega$, the lowest value of angular separation for which the two signals can be resolved or distinguished is called frequency resolution. The corresponding length of the window $w(t)$ is called time resolution, denoted by $\Delta\tau$. The frequency resolution $\Delta\omega$ and time resolution $\Delta\tau$ are inversely related by $\Delta\omega\Delta\tau \geq \frac{1}{2}$, which is a manifestation of the duality of STFT, inherited from the FT (Beecher 1988). The relation $\Delta\omega\Delta\tau \geq \frac{1}{2}$ is called the uncertainty principle, an analogy of a term used in quantum mechanics.

A special condition of equality in uncertainty relation, is satisfied by using a Gaussian window, but in practice are used a similar windows as the Hann, Hamming or Blackman-Harris. Consequently, the ability to timefrequency resolution of the STFT is fixed along the entire plane of time-frequency, a disadvantageous in signal analysis with respect to use of the Wavellet transform (Selin et al., 2007).

The essence of the STFT is to extract several frames of a signal using a window that moves over time. If the time window is sufficiently narrow, every frame extracted can be seen as stationary, so that the Fourier Transform (FT) can be used. The function of the STFT window is to use a part of the signal analysis and to ensure that the section is analyzed to be approximately stationary. To window perform its function, it is necessary to clarify two issues: the window type and size that should be used.

The spectrum of each window reveals parameters that characterize it for determined analysis. These parameters are the central lobe and side lobes (or main lobe and secondary lobes). The width of the central lobe is important to separate very close frequencies. The amplitude of the side lobes (relative to the amplitude of the central lobe) are important in controlling the degree of influence from a neighboring component in the other components of the signal (known as Spectral Leakage), and is related to abrupt variations in signal amplitude from start to finish of frame, and affect directly the formation of the spectrogram.

2.3.1 Spectrographic Image

An image is a picture, photograph or other form that gives a visual representation of an object or scene. In computer this representation is a two-dimensional array of numbers. Each number of the matrix corresponds to a small area, named pixel, and the values represent the number of each color image. This concept makes it possible to visualize an image as a three-dimensional graph commonly used to represent bivariate functions. The modulus squared of STFT of a signal $x(t)$ is called spectrogram of the signal.

The spectrogram represents an extension simple, but powerful, of the classical Fourier theory. In physical terms, the spectrogram provides a measure of the signal energy in the time-frequency plane. The spectrogram is an image that visually represents the frequency variation of a signal over time, and simultaneously gives a signal energy accurate and describes perfectly the details of the signal in question.

The color of a particular pattern in the spectrogram is indicative of the signal energy on that pattern. Generally, in a rating of notes, the harmonic sounds of a sonogram are required only when a hearing inspection is necessary. So it is convenient transform the sonogram in a pure JPEG image and calls it a spectrographic image for visual classification purposes. It is noteworthy that, in digital image processing (DIP), typically the coordinate system of an image is different from the graphics coordinate system, so it need a reversal in the ordinate axis so that the spectrogram can be interpreted coherently as a digital image.

2.4 Correlation

In obtaining two signals of a different experiment an issue that often arises is whether these two signals are correlated and the action signal triggers a response in the other signal. Depending on the simplicity of the signals is possible to find correlations between them by mere visual inspection.

However, for more complex signals (or noisy) detecting a correlation of these signals by visual inspection is impossible, then it becomes necessary to use mathematical and computational resources. The mathematical correlation between two signals $g(t)$ and $h(t)$ is defined as

$$g(t) \circ h(t) = \int_{-\infty}^{\infty} g(\tau) * w(t+\tau) dt \qquad (2)$$

The expression 2 multiplies two temporal series, being the second $h(t)$ shifted in time by an amount t (known as delay factor), and integrates the resulting signal. The inner product property suggests an interpretation of the mathematical correlation as a measure of similarity between two functions in relation to a number of relative displacements between them (Costa and Cesar Jr, 2000).

In this type of analysis the peaks are clearly identified and interpreted as follows: the correlation $(g \circ h)(t)$ on makes $h(t)$ glide along the function $g(t)$ and calculate the inner product for each of these situations, so that each intensity correlation gives an indication of similarity between functions (Costa and Cesar Jr, 2000).

In other words, the correlation allows searching for positions where two functions are more similar (in the broad sense of inner product).

An interesting application of correlation in PDI is a technique known as Template Matching, which seeks to locate a large image g the occurrence of similar objects to a small image h, the mold (or template). When calculating the correlation between the large picture and the mold, the positions at which the large image resembles the mold will have high correlation value. Thus, a search for the image of the extreme positions provides correlation, and the value of these extremes will provide the degree of similarity (Gonzalez et al., 2009).

The template matching has been widely applied in various areas. Its applications range from fingerprint recognition, face recognition, recognition of license plates, disease diagnosis from images, and calculations of similarities in the study of dialects Z.capensis (Avelino and Vielliard, 2004), which proves to be a very useful tool in pattern recognition. One may consider that inner products deployed by the correlation have sense only when the functions have their amplitudes properly normalized (Costa and Cesar Jr, 2000). So, an artifice often used to circumvent this problem is to perform the comparison via normalized correlation coefficients (NCC), which is defined by equation:

$$NCC = \frac{\sum_x \sum_y [g(x,y) - \bar{g}][h(x - x_i, y - y_i) - \bar{h}]}{\sqrt{\sum_x \sum_y [g(x,y) - \bar{g}]^2 \sum_x \sum_y [h(x - x_i, y - y_i) - \bar{h}]^2}}$$

Where $\bar{h}$ is the mean value of pixels in $h(x,y)$ (computed only once), $\bar{g}$ is the average value of $g(x,y)$ in the region coincident with the current location of h, and the sum is taken over the coordinates common to both g e h (Gonzalez et al., 2009). The $NCC(s,t)$ reaches values from -1 to 1, regardless of changes in the amplitude scale of $g(x,y)$ and $h(x,y)$. If the $NCC(s,t)$ is equal to -1 indicates that the mold found its negative.

2.5 Birdsong files on *www*

Some laboratories together with ecologists, has developed online services that offer a variety of tools for the examination and analysis of recordings made into the wild (Cugler et al., 2011; Towsey et al., 2012). The recordings of the *T. rufiventris* song were analogically equalized, digitized at a sampling rate of 44.1 *kHz*, edited and stored in a magneto-optic media. They were first digitized sounds of ASN and currently makes up one of the largest collections in the world of sounds made by animals such as birdsongs, the frogs croak or chirping insect, the Fonoteca Neotropical Jacques Vielliard (FNJV) of Institute of Biology, Unicamp (Cugler et al., 2011).

3 Material and methods

The sample is represented by professional quality recordings, most produced in Nagra E with ultradirecional microphone Senheiser MKH 816, others in DAT tape and some with ultradirecional microphone Senheiser ME 88. The recordings, performed in the wild, all properly identified as international standards (Kettle and Vielliard, 1991), are deposited in Neotropical Sound Archive (ASN), in University of Campinas (Unicamp). The international code reference contains the initials of the name of the collector, the number on the tape where the recording was made and the number of cut in chronological order.

In playback experiments, the chain of issuance was made with a Nagra E tape recorder connected to an amplifier 10 *W* and speaker customized with a frequency response: 100-86006 *Hz*, 2.5 *dB*, the sounds were issued at 90 *dB* SPL, measured at 1 *m* of speaker with a sound level meter equipped with a microphone Brüel & Kjaer 4126 of 1/2 *inch*.

The recorded are identified with the initials of phonoteca and their tumble number. The recorded chosen were: **FNJV5915**, **FNJV5932** and **FNJV5955** due of good technical quality and the specific characteristics of the song, as the amount and variety of notes in phrases. The software Matlab is capable of reading and playing files: wav, mono, in a sampling rate of 44.1 *kHz*, 16 *bit* and with a PCM compression format. Was developed a routine to read one phrase at a time and show its spectrogram. The sounds were visualized in spectrograms through a window $w(t)$ of Blackmann-Harris with 1024 bands, width of 100% and spectral energy logarithmic $75dB$. Later the spectrograms were transformed into figures.

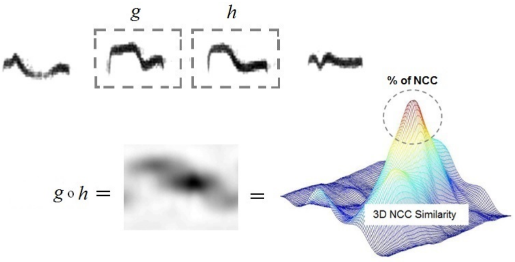

Figure 1: Similarity of two notes (g and h) by Template Matching. According to the human classification, the minimum value of degree of correlation to determine if two notes are the same is $NCC = 76\%$.

An image database was created from the pre-processed sounds. The routine developed in Matlab classifies notes of each phrase of the song of each individual by Template Matching technique. The calculations of NCC were performed by correlation property of the Fourier transform using a fast Fourier transform (FFT).

The notes of each phrase of the song were named by letters of the alphabet. The data of similarity between the notes were stored for possible structural analyzes of the repertoire. The criteria adopted for similarity and classification of notes in SITM was defined by degree of correlation (peak of NCC in percentage, figure 1). The minimum percentage of correlation that has agreement with Kappa test between the visual and auditory classification performed by ornithologists is $NCC = 76\%$. So, for that two notes are considered the same, the degree of correlation should be greater than 76%.

The first note that appears in the repertoire is classified as note **A**. If the second note has $NCC < 76\%$ compared to the first note (note **A**) it is classified as note **B**. If the third note has $NCC < 76\%$ compared to the first, but $NCC \leq 76\%$ compared to second note it is also classified as note B, and so on. The SITM performance was compared with a predefined classification (Da Silva et al., 2000) and second confidence intervals of statistical inference.

4 Results and discussion

Was analyzed the recorded of the birdsong of three individuals of the species T.rufiventris, the birdsong recorded **FNJV5915**, **FNJV5932** and **FNJV5955**.

The figures 2; 3 and 4; illustrates the spectrogram image of the typical T. *rufiventris* birdsong contained only in the recording **FNJV5915** and in Tables 1; 2 and 3; is presents the sequence of notes, represented by letters of the alphabet, in each phrase (P_i) of the recordings **FNJV** classified by SITM and confirmed visually by a trained human observer. In each test there was obtained a certain amount of hit, the total number of hit determines an average success within a certain number of samples, according to statistical inference.

In the recording **FNJV5915** contained five phases with seven different notes in the repertoire, and although the sequence is relatively simple, it is not sequentially progressive, and centralizes their scheme issued in note **A**, which initiates all phases. The SITM recognized correctly of 77% to 97% of the notes of the recording **FNJV5915**, with a confidence of 96%. The figure 2 shows the spectrogram of birdsong in **FNJV5915** and table 1 shows the SITM classification.

The figure 3 illustrates the spectrogram image of recording **FNJV5955**. The songbird of **FNJV5955** has four phrases and eleven types of notes, 74% to 97% of all notes were recognized and classified correctly by SITM with 96% of confidence. The table 2 shows the SITM classification of birsong in **FNJV5955**.

The repertoire restrained in the recording **FNJV5932** is only seven notes in nine phrases. The SITM recognized correctly 64% to 82% of all notes with a confidence 96%. The figure 4 illustrates the spectrogram image of recording **FNJV5932**. The table 3 shows the SITM classification of birsong in **FNJV5932**.

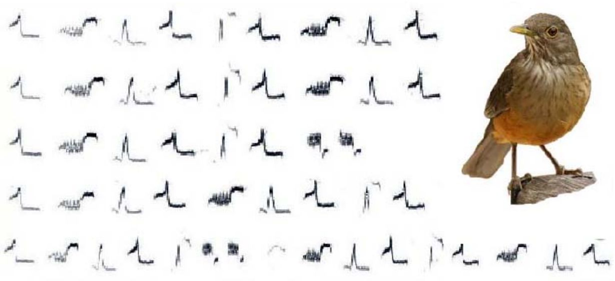

Figure 2: Phrases P_i of Rufous-bellied Thrush song in recording FNJV5915. The recognition of notes N_j is shown in table 1

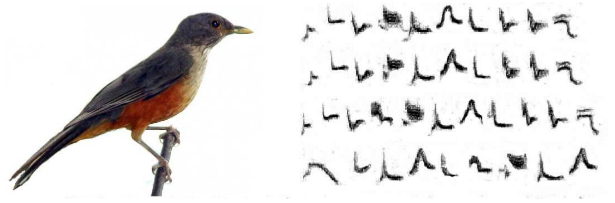

Figure 3: Phrases P_i of Rufous-bellied Thrush song in recording FNJV5955. The recognition of notes N_j is shown in table 2

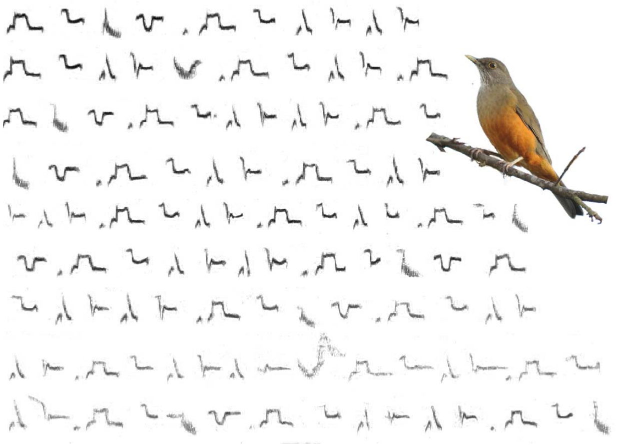

Figure 4: Phrases P_i of Rufous-bellied Thrush song in recording FNJV5932. The recognition of notes N_j is shown in table 3

There is a continuous debate in the literature about the appropriate use of alternative methods of comparing sounds. The study of Khanna et al. (Khanna et al., 1997) reveals the risk of running via comparisons with SPCC programs that easily make this type of analysis, as Signal or Canary, without examining the relevance of the spectrogram.

Table 1: Notes N_j of Phrase P_i of recording FNJV5915.

P_i/N_j	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}	N_{11}	N_{12}	N_{13}	N_{14}	N_{15}	N_{16}
P_1	A	B	C	A	D	A	B	C	A							
P_2	A	B	C	A	D	A	B	C	A							
P_3	A	B	C	A	D	A	D	D								
P_4	A	B	C	A	B	C	A	D	A	F						
P_5	A	B	C	A	D	E	E	G	B	C	A	D	A	B	C	A

Table 2: Notes N_j of Phrase P_i of recording FNJV5955.

P_i/N_j	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}	N_{11}
P_1	A	B	C	D	E	F	G	H	H	I	
P_2	A	B	C	D	E	F	G	H	H	I	
P_3	A	B	C	J	D	E	F	G	H	H	I
P_4	K	B	G	E	G	F	D	G	E		

Table 3: Notes N_j of Phrase P_i of recording FNJV5932.

P_i/N_j	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}	N_{11}	N_{12}	N_{13}	N_{14}	N_{15}
P_1	A	B	C	D	A	B	E	F	E	F					
P_2	A	B	E	F	G	A	B	E	F	A					
P_3	A	C	D	A	B	E	F	E	F	A	B				
P_4	C	D	A	B	E	F	E	F	A	B					
P_5	F	E	F	A	B	E	F	A	B	E	B	A	B	C	
P_6	D	A	B	E	F	E	F	A	B	C	D	A			
P_7	B	E	F	E	F	A	B	C	D	A	B	E	F		
P_8	E	F	A	B	E	F	E	F	G	A	B	E	F	A	B
P_9	E	F	A	B	C	D	A	B	E	F	E	F	A	B	C

An effective recognition can make SITM an efficient method of classification starting point to examine more bioacoustics chances. although there is a limitation in SITM when compared with other techniques similarity measures as used for Tchernichoviski et al. (Tchernichovski et al., 2000), the efficacy of the method yet can be improved by including more parameters in the similarity measure like details on the waveform of the note by three-dimensional graphical representation, using the usual tools in PDI (Brandes et al., 2006) such as mathematical morphology and shape and contour analysis as well as computational engines smarter (Deecke and Janik, 2006; Acevedo et al., 2009).

Many studies argue that the algorithms of supervised machine learning (Acevedo et al., 2009), such as linear discriminant analysis, decision trees, neural networks and Markov chains (Kogan and Margoliash, 1998) are the best choices for the automated identification of species, due to their high accuracy when compared with the human classification (Acevedo et al., 2009). But none of automated methods cited considers, for example, the importance of auditory perception of the species (Deecke and Janik, 2006).

The advantage of SITM relative to a trained observer human is the simultaneous processing of a large number of data and storage of information it provides facilities for further analysis. The image processing has an advantage over the sound processing with regard to pattern recognition.

5 Conclusion

Classification schemes may differ depending on contextual variable such as: habitat and effects associated with the propagation of sound; the identity of the singer and his social unit; The mechanism of sound production; how a receptor species categorizes the sound, or the performance of these sounds.

The classification is an important step in any study of behavior, and all classification problems require that decisions to be made by a human observer. Thus, the best approach arguably is one that combines the advantages of automatic methods with the ability of human observation. An integration of the complex human vision/auditory system with automated speed in handling a lot of data.

The present study can contribute significantly to the creation of such methodology for bioacoustics research, since it provides a means for at least partially of automatic categorization of vocalization related to a value set by parameters of human observations.

References

Brent Berlin and John P. O'Neill. The pervasiveness of onomatopeia in aguaruna and huambisa bird names. *J. Ethnobiol.*, 1(2):238–261, 1981.

Jacques Vielliard. *A ORNITOLOGIA NO BRASIL. Estado atual das pesquisas em bioacústica suas sua contribuição para o estudo e a proteção das aves do Brasil.* UERJ, Rio de Janeiro, 1st edition, 2000.

Jacques Vielliard. Uso da bioacústica na observação de aves. In *In: Coelho EP. (Ed), II Enc Nac Anilhad Aves. Rio de Janeiro: UFRJ*, pages 98–121, 1987.

David B Pisoni, Beth G Greene, and Thomas D Carrell. Identification of visual displays of speech: comparisons of naive and trained observers. *The Journal of the Acoustical Society of America*, 73:S102, 1983.

Christopher W Clark, Peter Marler, and Kim Beeman. Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology*, 76(2): 101–115, 1987.

Zhixin Chen and Robert C Maher. Semi-automatic classification of bird vocalizations using spectral peak tracks. *The Journal of the Acoustical Society of America*, 120:2974, 2006.

Márcio F Avelino and Jacques ME Vielliard. Comparative analysis of the song of the rufous-collared sparrow zonotrichia capensis (emberizidae) between campinas and botucatu, são paulo state, brazil. *Anais da Academia Brasileira de Ciências*, 76(2):345–349, 2004.

Hernán Fandiño-Mariño and Jacques ME Vielliard. Complex communication signals: the case of the blue-black grassquit volatinia jacarina (aves, emberizidae) song. part i-a structural analysis. *Anais da Academia Brasileira de Ciências*, 76(2):325–334, 2004.

Erich D Jarvis, Sidarta Ribeiro, Maria Luisa Da Silva, Dora Ventura, Jacques Vielliard, and Claudio V Mello. Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature*, 406(6796):628–632, 2000.

Maria Luisa Da Silva, Jose Roberto C Piqueira, and Jacques ME Vielliard. Using shannon entropy on measuring the individual variability in the rufous-bellied thrush turdus rufiventris vocal communication. *Journal of Theoretical Biology*, 207(1):57–64, 2000.

Amanda Freitas Valentim, Marcela Guimarães Côrtes, and Ana Cristina Côrtes Gama. Spectrographic analysis of the voice: effect of visual training on the reliability of evaluation. *Revista da Sociedade Brasileira de Fonoaudiologia*, 15(3):335–342, 2010.

Ana Cristina Côrtes Gama, Luiza Lara Marques Santos, Natália Aparecida Sanches, Marcela Guimarães Côrtes, and Iara Barreto Bassi. Studying the effect of spectrogram visual support of in the auditory-perceptive voice evaluation reliability. *Revista CEFAC*, 13(2):314–321, 2011.

Leo Cohen. *Time Frequency Analysis: Theory and Applications*. Prentice Hall PTR, 2st edition, 1995.

Arja Selin, Jari Turunen, and Juha T Tanttu. Wavelets in recognition of bird sounds. *EURASIP Journal on Applied Signal Processing*, 2007(1):141–141, 2007.

Luciano da Fontoura Da Costa and Roberto Marcondes Cesar Jr. *Shape analysis and classification: theory and practice*. CRC Press, Inc., 2000.

Rafael C Gonzalez, Richard E Woods, and Steven L Eddins. *Digital image processing using MATLAB*, volume 2. Gatesmark Publishing Knoxville, 2009.

Daniel Cintra Cugler, Claudia Bauzer Medeiros, and Luıs Felipe Toledo. Managing animal sounds-some challenges and research directions. In *Proceedings V eScience Workshop-XXXI Brazilian Computer Society Conference*, 2011.

Michael Towsey, Birgit Planitz, Alfredo Nantes, Jason Wimmer, and Paul Roe. A toolbox for animal call recognition. *Bioacoustics*, 21(2):107–125, 2012.

Ron Kettle and Jacques Vielliard. Documentation standards for wildlife sound recordings. *Bioacoustics*, 3(3): 235–238, 1991.

Khanna, Gaunt, and McCallum. Digital spectrographic cross-correlation: tests of sensitivity. *Bioacoustics*, 7(3): 209–234, 1997.

Ofer Tchernichovski, Fernando Nottebohm, Ching Elizabeth Ho, Bijan Pesaran, and Partha Pratim Mitra. A procedure for an automated measurement of song similarity. *Animal Behaviour*, 59(6):1167–1176, 2000.

Scott Brandes, Piotr Naskrecki, and Harold Figueroa. Using image processing to detect and classify narrowband cricket and frog calls. *The Journal of the Acoustical Society of America*, 120:2950, 2006.

Volker B Deecke and Vincent M Janik. Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. *The Journal of the Acoustical Society of America*, 119:645, 2006.

Miguel A Acevedo, Carlos J Corrada-Bravo, Héctor Corrada-Bravo, Luis J Villanueva-Rivera, and T Mitchell Aide. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4):206–214, 2009.

Joseph A Kogan and Daniel Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: A comparative study. *The Journal of the Acoustical Society of America*, 103:2185, 1998.