Citation:

Sandoval L, Barrantes G, Wilson DR (2019) Conceptual and statistical problems with the use of the Shannon-Weiner entropy index in bioacoustic analyses. *Bioacoustics*, 28: 297–311. doi: 10.1080/09524622.2018.1443286

Conceptual and statistical problems with the use of the Shannon-Weiner entropy index in bioacoustic analyses

Luis Sandoval[a], Gilbert Barrantes [a] and David R. Wilson[b]

[a]Escuela de Biología, Universidad de Costa Rica, San Pedro, San José, Costa Rica CP-2090; [b]Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

Corresponding author: Luis Sandoval. Escuela de Biología, Universidad de Costa Rica, San Pedro, San José, Costa Rica CP-2090. e-mail: biosandoval@hotmail.com

Conceptual and statistical problems with the use of the Shannon-Weiner entropy index in bioacoustic analyses

Information theory and its indices were developed for human communication to predict the amount of information transferred in a message. One such index, the Shannon-Weiner index (SWI), has often been used to analyze information from other fields in which its application may not be appropriate. In ecoacoustics, the SWI is used to compare acoustic diversity (*i.e.,* a measure derived by integrating the richness and abundance of animal sounds) among locations. In animal communication, the SWI is used to quantify repertoire complexity (*i.e.*, a measure derived by integrating the number and abundance of sound types produced by individuals or species) as an approach to understanding signal evolution. We discuss problems associated with using the SWI in ecoacoustics and animal communication. Specifically, we discuss conceptual and statistical problems associated with the SWI, and then illustrate these problems using hypothetical data. In ecoacoustics, the SWI's assumptions of random variables and independent samples are often violated. In animal communication, the SWI fails to distinguish among repertoires in which the number of sound types and the abundance of each sound type differ. We also show that other methods do capture these differences. We conclude that the SWI does not adequately represent acoustic diversity or repertoire complexity due to the multiple conceptual and statistical issues associated with its use. We recommend other analytical methods to more fully describe these biological systems, including goodness of fit, Morisita similarity index, and Markov chain analysis. These methods provide more information for future comparisons, and permit researchers to test hypotheses more directly.

Keywords: ecoacoustics; bioacoustics; repertoire; Shannon-Weiner entropy index; statistical analysis; soundscape

Introduction

The Shannon-Weiner entropy index (SWI) was developed to measure the amount of information transferred in a message over telephone lines (Shannon and Weaver 1949). This index estimates the uncertainty in the information code of a message (Pielou 1966; Krebs 1999), but does not estimate the number of information codes (Jost 2006), the specific codes included in the message, or the order in which the codes are produced (Palmero et al. 2014). Since its origin, the SWI, which is also called first-order entropy (McCowan et al. 1999, 2002), has been used extensively in community ecology and population genetics (*e.g.*, Meirmans and Van Tienderen 2004; Forster et al. 2006; Sherwin 2010; Peakall and Smouse 2012). However, its use in those fields has been criticized because it: (1) condenses two unrelated variables into a single metric (*e.g.*, species composition and the abundance of individuals in each species; Allen et al. 2009; Barrantes and Sandoval 2009); (2) is very sensitive to small samples (Peet 1975; Green 1979); and (3) does not adequately reflect rare species or alleles (Chao and Shen 2003). More recently, and despite its inappropriate use in other fields, the SWI has also been applied to the fields of ecoacoustics and animal communication.

In the field of ecoacoustics (Sueur and Farina 2015), and similar areas such as soundscape ecology (Farina et al. 2011; Krause 2016), the use of the SWI is becoming widespread (Pieretti et al. 2011; Depraetere et al. 2012). The main goal of ecoacoustics, as stated by Sueur and Farina (2015), is to be an "*applied discipline that studies sound along a broad range of spatial and temporal scales in order to tackle biodiversity and other ecological questions*". Within this broad goal, a common approach is to relate the acoustic environment to species richness and the abundance of individuals within each species

(Pijanowski et al. 2011; Depraetere et al. 2012; Sueur et al. 2012). For example, Depraetere et al. (2012) tried to determine the relation between sound recordings and species richness and abundance. They asked: "(*i*) *do the indices match with results provided by a classical bird inventory?... and (iii) could the indices highlight expected biodiversity differences between different habitats?*" Research in this field uses autonomous audio recorders to monitor the acoustic environment over long periods of time (*e.g.*, hundreds or thousands of recording hours; Blumstein et al. 2011; Mennill et al. 2012; Sueur et al. 2012). The large acoustic datasets are then analyzed using different data extraction procedures that usually involve automatic detection of animal signals (Sueur et al. 2012). After the data are extracted, some studies estimate the diversity of the acoustic environment (i.e., the number of species detected and the frequency of occurrence of each species' signals) by integrating all of the data into a single measure using information theory indices, such as the SWI (Sueur et al. 2012; Gasc et al. 2013). However, the SWI does not directly reflect species richness, the abundance of individuals within each species, or species composition, so a large portion of the original information is lost.

Animal communication is another field in which the use of the SWI is becoming widespread. Over the last seven decades, the field has amassed large literatures that describe how information is encoded in acoustic signals through structural variation, sequence level variation (*e.g.,* number of signals or signaling rate), and syntactical rules (Gerhardt and Huber 2002; Marler 2004). Much of this effort has centered on identifying the fundamental units of communication (Bradbury and Vehrencamp 2011). Traditionally, information encoding mechanisms were analyzed by identifying and counting the number of different sound units (*e.g.,* syllables or elements) produced by each individual animal or species, and by then examining the order in which those units were produced (Botero et al.

2008; Vargas-Castro et al. 2012; Sandoval et al. 2014). Recently, the SWI has replaced these more traditional measures (McCowan et al. 1999; Aubin et al. 2004; Palmero et al. 2014). In spectacled warblers (*Sylvia conspicillata*) and bottlenose dolphins (*Tursiops truncatus*), for example, the SWI was used to calculate the entropy or complexity of their communication systems. The index incorporated the occurrence of different types of syllables, as well as the probabilities of their occurrence (McCowan et al. 1999; Palmero et al. 2014).

Another problem in both fields is that multiple terms are used interchangeably (*e.g.*, diversity, complexity, entropy, richness, composition, and randomness). This results in a confusing terminology that reduces the comparative scope of these studies. For instance, the SWI is often used to quantify different things, such as diversity, complexity, and entropy. As an example, 'diversity' is often labeled as 'complexity' or 'entropy' in sound analyses, though each of these terms has a different meaning. Diversity is an integrated measure of the number of sound types and the abundance of each sound type that are produced by an individual, population, or species. Song complexity, which sometimes is used interchangeably with richness and composition, is an integrated measure of the number of different elements or syllables produced in each song (Buchanan and Catchpole 1997; Palmero et al. 2014) and the entropy order and versatility of the internal song structure (Hamao 2008). Song entropy, which sometimes is interchanged for randomness, is a measurement of song organization (first-order of entropy according to McCowan et al. 1999; Palmero et al. 2014), where higher SWI values indicate higher entropy and a more even distribution of sounds among sound types.

Our objectives in this paper are: (1) to describe general conceptual and statistical problems inherent to the use of the SWI in ecoacoustics; and (2) to highlight the disadvantages of using information theory indices in studies of animal communication by analyzing hypothetical acoustic repertoires using the SWI and other alternative statistical techniques.

Case Studies

Ecoacoustics

In ecoacoustics, investigators use the SWI to estimate biological diversity because the SWI combines sound richness and the abundance of sounds in each sound type into a single metric. Sound richness is determined by the number of unique sound types or the number of unique species that are detected in audio recordings, whereas sound abundance is determined by how often each sound type or species is detected over time.

A fundamental assumption of the SWI is that it measures the uncertainty of occurrence of a random variable, such as the probability that a particular letter will appear next in a string of text (Pielou 1966). However, animal sounds are not produced at random (Staicer et al. 1996). Rather, they exhibit diel and seasonal patterns (Staicer et al. 1996), respond predictably to non-random biotic and abiotic interference (Slabbekoorn 2004), and change in response to non-random intraspecific and interspecific social interactions (Bradbury and Vehrencamp 2011). This non-randomness in the context of ecoacoustics violates a fundamental assumption of the SWI (Pielou 1966; Krebs 1999).

The SWI includes in its formula the proportional contribution of each sound type to the total number of sounds in the sample:

$$\mathrm{SWI} = -\sum_{i=1}^{s} (p_i)(\log p_i)$$

where *s* is the number of sound types or species, and p_i is the proportion of the total sample belonging to *i*th sound type or species. In this formula, p_i is multiplied by $\log p_i$ because, in order to estimate the total complexity or diversity of sound types or species, it is necessary to average the potential contributions of each sound type or species (Ulanowicz 1997). A second problem with the SWI is that its value increases in a nonlinear fashion as the number of sound types or species in the sample increases (Wolda 1981; Krebs 1999; Jost 2006). As an example, Jost (2006) showed that a community with eight equally common species had a SWI value of $H' = 2.0$, whereas a community with 16 equally common species had a SWI value of $H' = 3.0$. In this example, the community with 16 equally common species has twice as many species and twice as much diversity as the community with eight equally common species, but the ratio of the two SWI values is only 3:2. Additionally, a common practice when using the SWI is to covert H' values into evenness values using the $e^{H'}$ formula. This formula provides the total number of species, assuming equal abundances, based on the H' value. But, for this example, the formula $e^{H'}$ yields 7.4 species when $H' = 2.0$ and 20 species when $H' = 3$. In both cases, the estimated number of species differs from the real number (8 and 16 equally abundant species, respectively). This nonlinear relationship between diversity and SWI values reduces the utility of the SWI in comparative studies because the SWI values are not directly proportional to species richness, species abundance, or diversity.

A third problem is that communities with different values of richness and abundance can produce the same SWI value. Consider two communities that each contain 80 individuals. One community includes ten individuals from each of eight species, whereas the other includes 35 individuals of one species, six individuals from each of two species, five individuals from each of five species, and one individual from each of eight species. Despite their obvious differences in species richness and abundance, these two communities yield the same SWI value (Table 1). This is because the SWI penalizes rare species (Chao and Shen 2003) and does not fully capture other important aspects of a community, such as richness and abundance (Allen et al. 2009; Barrantes and Sandoval 2009), thus limiting its utility in comparing communities. In contrast, the Morisita Index considers species abundance and richness, and thus can distinguish among communities with similar diversity, but which differ in these other community metrics. In this example, the similarity of the two communities, according to the Morisita Index, is 66%, a value that better reflects the differences in the species richness and abundance of the two communities. Diversity (as defined by SWI) is only one parameter of a community, but, on its own, often provides little information. Thus, communities should be characterized by direct measures of abundance, richness, and composition, in addition to diversity or diversity indices, such as the Morisita index (used here), NESS (normalized expected species shared) index (Grassle and Smith 1976), their generalized versions (Chao et al. 2008), and Bray-Curtis (Bloom 1981) that preserve variation in each of these fundamental characteristics. It is important to mention here (although it is not the goal of this paper) that the Morisita index has been criticized because its calculation is affected by species abundance (see Ricklefs and Lau 1980; Bloom 1981; Chao et al. 2006, 2008 for discussion about this topic). However, this characteristic makes this index robust when individual repertoires are not completely sampled because the most common sounds are present in the sampling effort (Chao et al. 2006).

A fourth problem when using the SWI in ecoacoustic studies is that it does not provide an error estimation (*e.g.,* log-likelihood or residual sum of squares). Rather, the SWI is a single value derived from the number of sound types or species and the abundance of each sound type or species at a given location. Therefore, SWI values cannot be adjusted to a particular probability distribution, which reduces their utility in comparative studies. The lack of an error term also makes it difficult to calculate an effect size, which is the basic measurement used in meta-analysis (Arnqvist and Wooster 1995). Some investigators suggest that multiple recordings be obtained from the same location (or that a single recording be subdivided into multiple smaller recordings), and that the jackknife or bootstrapping approach be used to estimate confidence intervals around the mean SWI value (Adams and McCune 1979). Although confidence intervals are useful, they are not the same as measures of error because confidence intervals are based on observed variation, whereas error terms are based on how much observed values differ from expected values derived from a theoretical distribution.

Animal communication

We analyzed repertoire complexity (*i.e.,* number of song types and abundance of each song type per individual) using four simulated data sets. We chose this method because it provides precise control over sample size, repertoire complexity, and repertoire size.

We created the four data sets such that each of them contained different repertoire complexities: (1) a data set in which individuals' repertoires contained the same two song types at various ratios (Table S1), (2) a data set in which individuals' repertoires contained the same eight song types at various ratios (Table S2), (3) a data set in which individuals' repertoires contained the same 20 song types at various ratios (Table S3), and (4) a data set in which individuals' repertoires contained between 2 and 16 song types, and in which the number of songs of each song type varied among individuals (Table S4). This last data set is representative of several avian species in which conspecifics have different repertoire sizes (*e.g.*, Botero et al. 2008; Sandoval et al. 2014). The first three data sets contained 100 songs from each of 20 individuals, and the fourth data set contained 100 songs from each of 30 individuals (Tables S1-S4). These data sets were selected to illustrate species with small, medium, and large song repertoires, and to illustrate the inability of the SWI to distinguish among individuals with different pattern of song production.

In each of the first three data sets, we divided the 20 individuals into two groups of 10. For the first group (individuals 1-10; Tables S1-S3), we controlled the distributions of songs among song types, so that they ranged from an individual having all songs represented in the same proportion (*i.e.,* individual 1; Tables S1-S3) to an individual having an extremely uneven distribution of songs among song types (*i.e.,* individual 2). For the second group (individuals 11-20; Tables S1-S3), we used the 'random' function in Excel (version 2007 for Windows; Microsoft Corporation, Redmond, WA, USA) to randomly create each individual's distribution of songs among the available song types. In the fourth data set, we varied the number of song types included in each individual's repertoire from 2 to 16 (Table S4). We also created distributions in which songs were distributed evenly among song types for 15 individuals (ev2-ev16; Table S4), and in which they were distributed extremely unevenly among song types for the other 15 individuals (sk2-sk16). This fourth data set was selected to illustrate how differences in song richness and abundance can produce similar measures of diversity, complexity, or randomness, as quantified by the SWI. In all four data sets, we assigned songs to song types in a random order, even though their probabilities of being assigned to each song type were often quite different.

Following the approach used in recent studies of repertoire complexity (Aubin et al. 2004; Kershenbaum 2013; Palmero et al. 2014), we used the SWI to compare repertoire complexities among individuals from the same data set. We calculated the SWI value (H') for each individual based on the natural logarithm, and estimated its 95% confidence interval using a bootstrap approach with 9999 random permutations. For each permutation, one of the 100 songs of a given individual was selected at random and excluded before re-calculating H'.

SWI values are difficult to interpret because they do not denote the original biological units that were used to create them. We therefore exponentiated our SWI values by calculating e to the power of H' to obtain biologically meaningful units (in this case, song types), as recommended by Jost (2006). However, because most studies present only the original SWI values, we report both the original (H') and the converted values (H_c').

In addition to the SWI, we used three other statistical tests to compare individuals within a data set and to show that these methods provide a better characterization of interindividual differences than the SWI alone. First, for data sets 1–3, we used a chi-square goodness-of-fit test to determine if the distribution of song types varied between the 20 individuals. For this test, we expect that individuals that have a similar abundance of each song type would also have similar SWI values. Second, we used a Morisita index of similarity to determine whether song repertoires (richness and abundance) were similar (values near 1) or different (values near 0) among individuals. The Morisita index incorporates repertoire size (richness) and the abundance of each song type, and its results are presented using a cluster analysis. We tested for differences among the clusters of individuals using one-way analysis of similarity (ANOSIM). Statistical differences obtained with this analysis indicate that richness and abundance differ between groups of individuals, and that individuals within groups have similar composition. In ANOSIM, there is no set rule for defining groups, rather, groups are usually defined a priori based on knowledge on the working system (*e.g.,* individuals 1–5 are from one population, while individuals 6–12 are from a different population, and so on). In our datasets, there was no a priori knowledge about grouping structure, so we determined the grouping structure through post hoc inspection of the cluster trees. Third, we conducted a Markov chain analysis for each individual in the second data set to illustrate the potential use of this technique to describe repertoire entropy (sometimes also called repertoire randomness) characteristics that can also be analyzed using a second-order SWI (*e.g.*, McCowan et al. 1999; Dayou et al. 2011; Palmero et al. 2014). Markov chain analysis reports the probability that the sample was drawn from an individual in which all possible transitions between song types are equally probable (*i.e.,* the choice of song type does not depend on which song type was sung last). All statistical analyses were conducted using PAST 2.17 (Hammer et al. 2001).

Results first scenario

In this case, the entropy of the repertoire ranged from H' = 0.06 (H_c’ = 1.05 song types) for individual i18 to H' = 0.69 (H_c’ = 2.00 song types) for individual i1 (Fig. 1a). Overall, the distribution of each individual’s 100 songs between the two song types differed significantly among the 20 individuals ($\chi^2 = 566.77$, df = 19, $p < 0.001$). Individuals *i6* and *i7* had exactly the same SWI values for their repertoires (Fig. 1a), yet they differed the most in the proportion of each song type according to the Morisita index of similarity (Fig. 1b, S1). The cluster tree showed three groups of individuals (Fig. 1b), with individuals in each cluster being significantly more similar to each other than to individuals from other clusters (ANOSIM using Morisita scores: R = 0.85, p = 0.001).

Results second scenario

For this scenario, the entropy of the repertoire ranged from H' = 0.39 (H_c' = 1.48 song types) for individual i10 to H' = 2.08 (H_c' = 8.00 song types) for individual i1 (Fig. 2a). Overall, the distribution of each individual's 100 songs among the 8 song types differed significantly among the 20 individuals ($\chi^2 = 874.42$, df = 133, $p < 0.001$). Individuals *i3*, *i4,* and *i5* had exactly the same SWI values for their repertoires (Fig. 2a), yet the abundance of each song type in their repertoires varied by up to 20% according to the Morisita index of similarity (Fig. 2b). For individuals whose repertoires were created randomly, and whose entropy values were similar (*i.e.,* had overlapping 95% confidence intervals in Fig. 2a), repertoire similarities varied from only 2 to 12% according to the Morisita index (Fig. 2b, S2). The cluster tree showed four groups of individuals (Fig. 2b), with individuals in each group being significantly more similar to each other than to individuals from other groups (ANOSIM using Morisita scores: R = 0.74, $p < 0.001$).

The SWI quantifies the randomness of the distribution of items (e.g., songs) among categories (*e.g.,* song types). It was not designed to quantify the randomness of the order in which items from different categories appear (e.g., AABB versus ABAB), though it has often been used for this purpose. Our Markov chain analysis showed that 11 individuals produced songs in a random order (Table 1; all $p > 0.1$), and that 9 individuals did not (Table 1; all $p < 0.001$). Furthermore, some individuals that produced their song repertoire in a random order had SWI values that were indistinguishable from those of individuals that produced their song repertoire in a non-random order. For example, individuals *i14* and *i16* had the same SWI values, yet *i14* produced its songs in a non-random order, while *i16* produced its songs in a random order (Table 1; Fig. 2a). Thus, the SWI does not reliably distinguish individuals that produce their songs in a random order from those that produce their songs in a non-random order.

Results third scenario

The entropy of the repertoire ranged from H' = 1.05 (H$_c$' = 2.86 song types) for individual *i10* to H' = 3.00 (H$_c$' = 20.01 song types) for individual *i1* (Fig. 3a). The distribution of each individual's 100 songs among the 20 song types differed significantly among the 20 individuals (χ^2 = 1133.60, df = 361, $p < 0.001$). Individual *i5* and *i3* had the same entropy value (*i.e.,* H' = 2.54; H$_c$' = 12.63 song types; Fig. 3a), yet the abundance of each song type in their repertoires was quite different (approximately 40% according to the Morisita index of similarity; Fig. 3b, S3). In contrast, individuals *i5* (H' = 2.53; H$_c$' = 12.55 song types) and *i9* (H' = 2.15; H$_c$' = 8.58 song types) had markedly different entropy values (Fig. 3a), yet the abundance of each song type in their repertoires was more similar (30% according to the Morisita index of similarity; Fig. 3b). The cluster tree showed three groups of individuals (Fig. 3b), with individuals in each group being more similar to each other than to individuals from other groups (ANOSIM using Morisita scores: R = 0.88, $p < 0.001$).

Results fourth scenario

Among the 15 individuals that had songs assigned to song types according to a skewed distribution, the entropy of the repertoire ranged from H' = 0.06 (H$_c$' = 1.06 song types) for individual *sk*2 to H' = 0.83 (H$_c$' = 2.29 song types) for individual *sk*16 (Fig. 4a). Among the 15 individuals that had their songs distributed evenly among song types, repertoire entropy ranged from H' = 0.69 (H$_c$' = 2.00 song types) for individual *ev*2 to H' = 2.77 (H$_c$' = 15.96 song types) for individual *e*16 (Fig. 4a). Individuals with nine or more song types in their repertoire and a skewed distribution of songs had entropy values that were statistically indistinguishable from those of individual *ev*2 (based on overlapping 95% confidence intervals), whose songs were distributed evenly among only 2 song types (Fig. 4a, S4). According to the Morisita index of similarity, increasing repertoire size had the smallest effect on repertoire similarity when repertoires were large and songs were evenly distributed among song types (Fig. 4b). The cluster tree showed three groups of individuals (Fig. 4b), with individuals from the same group being significantly more similar to each other than to individuals from other groups (ANOSIM using Morisita scores: R = 0.16, p = 0.04).

DISCUSSION

Ecoacoustics is a developing field that bridges diverse areas of investigation, including biodiversity, urban development, changes in landuse (*e.g.,* mining, forestry, and agriculture), and conservation (Truax and Barrett 2011; Farina and Pieretti 2012; Sueur et al. 2012). As a complex and flourishing field, a diverse set of methods has been developed to compare biological communities based on the sounds recorded at different locations. One method that has become popular for analyzing those data in recent years is the SWI. However, as we have argued here, the SWI has several inherent problems that undermine its validity in studies of ecoacoustics. For example, when presented on its own, the SWI fails to adequately describe biological communities because it does not consider the specific species in a community, but, rather, reduces the number of species and the number of individuals in each species to a single number. As a result, communities with different species compositions, different number of species, and different distributions of individuals among species can all have the same SWI value, despite their obvious differences. The absence of error terms around the SWI values precludes the calculation of effect size, which is the basic measurement used in meta-analysis (Arnqvist and Wooster 1995). This makes it difficult to include results of the SWI in meta-analyses, which are very valuable for evaluating general patterns and for resolving the complex interactions that occur among animal species and other abiotic factors (*e.g.,* noise, habitat structures, or urban development). Therefore, in ecoacoustics, we encourage researchers to use or develop analyses that provide error terms, such as likelihood or odd ratios.

Several descriptive and statistical methods can be used to analyze the complexity or diversity of an animal's acoustic repertoire (Botero et al. 2008; Sandoval et al. 2014). The chosen method depends on the question to be answered and the complexity or diversity of the repertoire in terms of syntax, number of sound types, and the distribution of sounds among sound types. The SWI conveniently reduces each individual's repertoire or the repertoire of the entire community to a single value, but that value does not indicate the specific sound types in the repertoire, the sound type richness, the distribution of sounds among sound types, or the order in which sound types are produced. Therefore, when used by itself, the SWI may not reveal fundamental differences among individuals or communities. For example, it would not distinguish between an individual that sings song types a and b at a 1:3 ratio and an individual that sings those same song types at a 3:1 ratio. Furthermore, differences in SWI values can be difficult to interpret because they could simply reflect the random error created by incomplete sampling of each individual's repertoire; meaningful comparisons can only be made by computing and comparing confidence intervals for each SWI value, as advocated by Adams and McCune (1979) and as demonstrated in our examples (Figs. 1-4). However, even this method is not ideal because the 95% confidence intervals are very wide, especially when song types are distributed unevenly among song types (Fig. 4a). The confidence intervals also tend to increase with increasing repertoire size (Fig. 4a). Consequently, there is a low probability that the SWI will distinguish between repertoires of unequal complexity, especially when songs are distributed unevenly among song types, and especially for individuals with larger repertoires.

We recommend that a combination of techniques be used when describing and comparing biological communities in studies of ecoacoustics, or vocal repertoires in studies of animal communication. Beginning by reporting the specific sound types or species detected, the number of sound types or species detected, and the population-level distribution of sounds or individuals among sound types/species. Then run a chi-square goodness of fit test to test if the distribution of sounds/individuals among sound types/species varies among individuals/locations. If it does, then a Morisita index can be used to quantify similarity among individuals/locations, and an ANOSIM can be used to test for differences among any set of groups that were known a priori (*e.g.,* two different populations of the same species). For studies of animal communication, a Markov chain analysis can also be used to test the randomness or complexity of songs. This could be conducted on the entire population, or, if the contingency table analysis was significant, then perhaps on each individual separately.

In conclusion, the SWI provides only a poor representation of complexity inherent to the fields of ecoacoustics and animal communication. In ecoacoustics, it is important to preserve information about the number of species, species composition, and the distribution of individuals among species, since these parameters result from different and unrelated processes (Barrantes and Sandoval 2009). Yet, communities that differ greatly in these parameters can yield identical SWI values. This issue is especially important when dealing with changes in species composition or conservation because not all species have the same ecological role or the same conservation problems. In animal communication, the SWI is also an oversimplification of biological complexity because it reduces the complexity of an individual's vocal repertoire to a single value that does not reliably reflect repertoire size, repertoire composition, the distribution of sounds among sound types, or the animal's syntactical rules. Other statistical methods, such as the contingency table analysis, Morisita index of similarity, and Markov chain analysis, are more informative and more conducive for comparisons among studies.

Acknowledgements, we thank B. McCowan and Robert E. Ulanowicz for valuable comments that improved a previous draft of this paper. LS and GB also thanks to Vicerrectoría de Investigacíon, Universidad de Costa Rica all the support for this project.

References

Adams JE, McCune ED. 1979. Application of the generalized jack-knife to Shannon's measure of information used as an index of diversity. In: Grassle KF, Patil GP, Smith W, Taillie C, editors. Ecological diversity in theory and practice. Maryland: International Cooperative Publishing House; p. 117–131.

Allen B, Kon M, Bar-Yam Y. 2009. A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. American Naturalist. 174: 236–243.

Arnqvist G, Wooster D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution. 10:236–240.

Aubin T, Mathevon N, Silva MLD, Vielliard JM, Sebe F. 2004. How a simple and stereotyped acoustic signal transmits individual information: the song of the White-browed Warbler *Basileuterus leucoblepharus*. Anais da Academia brasileira de Ciencias. 76:335–344.

Barrantes G, Sandoval L. 2009. Conceptual and statistical problems associated with the use of diversity indices in ecology. Revista de Biología Tropical. 57:451–460.

Bloom SA. 1981. Similarity indices in community studies: potential pitfalls. Marine Ecology Progress Series. 5:125–128.

Blumstein DT, Mennill DJ, Clemins P, Girod L, Yao K, Patricelli G, Deppe JL, Krakauer AH, Clark C, Cortopassi KA, Hanser SF, McCowan B, Ali AM, Kirscehl ANG. 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations, and prospectus. Journal of Applied Ecology. 48:758–767.

Botero CA, Mudge AE, Koltz AM, Hochachka WM, Vehrencamp SL. 2008. How reliable are the methods for estimating repertoire size? Ethology. 114:1227–1238.

Bradbury JW, Vehrencamp SL. 2011. Principles of animal communication: 2nd edn. Sunderland, MA: Sinauer Associates.

Buchanan KL, Catchpole CK. 1997. Female choice in the sedge warbler *Acrocephalus schoenobaenus* multiple cues from song and territory quality. Proceedings of the Royal Society of London B. 264:521–526.

Chao A, Shen TJ. 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. Environmental and Ecological Statistics. 10:429–443.

Chao A, Chazdon RL, Colwell RK, Shen, TJ. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics. 62:361–371.

Chao, A., Jost L, Chiang SC, Jiang YH, Chazdon RL. 2008. A two-stage probabilistic approach to multiple-community similarity indices. Biometrics. 64:1178–1186.

Dayou J, Han NC, Mun HC, Ahmad AH. 2011. Classification and identification of frog sound based on entropy approach. IPCBEE. 3:184–187.

Depraetere M, Pavoine S, Jiguet F, Gasc A, Duvail S, Sueur J. 2012. Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. Ecological Indicators. 13:46–54.

Farina A, Pieretti N. 2012. The soundscape ecology: a new frontier of landscape research and its application to islands and coastal systems. Journal of Marine and Island Cultures. 1:21–26.

Farina A, Pieretti N, Piccioli L. 2011. The soundscape methodology for long–term bird monitoring: a Mediterranean Europe case–study. Ecological Informatics. 6:354–363.

Forster RM, Créach V, Sabbe K, Vyverman W, Stal LJ. 2006. Biodiversity–ecosystem function relationship in microphytobenthic diatoms of the Westerschelde estuary. Marine Ecology Progress Series. 311:191–201.

Gasc A, Sueur J, Jiguet F, Devictor V, Grandcolas P, Burrow C, Depraetere M, Pavoine S. 2013. Assessing biodiversity with sound: do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? Ecological Indicators. 25:279–287.

Gerhardt HC, Huber F. 2002. Acoustic communication in insects and anurans: common problems and diverse solutions. Chicago, Il: University of Chicago Press.

Green RH. 1979. Sampling design and statistical methods for environmental biologists. New York, NY: Wiley.

Hamao S. 2008. Syntactical complexity of songs in the black–browed reed warbler *Acrocephalus bistrigiceps*. Ornithological Science. 7:173–177.

Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica. 4:9. http://palaeo–electronica.org/2001_1/past/issue1_01.htm

Jost L. 2006. Entropy and diversity. Oikos. 113:363–375.

Kershenbaum A. 2013. Entropy rate as a measure of animal vocal complexity. Bioacoustics. 23:195–208.

Krebs CJ. 1999. Ecological methodology., Menlo Park, CA: Benjamin Cummings.

Krause B. 2016. Wild soundscape, discovering the voice of the natural world. 2nd. edn. New Haven, CT: Yale University Press.

Marler P. 2004. Science and birdsong: the good old days. In: Marler P, Slabbekoorn H, editors. Nature´s music, the science of bird song. San Diego: Elsevier Academic Press; p. 1–38.

McCowan B, Hanser SF, Doyle LR. 1999. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. Animal Behaviour. 57:409–419.

McCowan B, Doyle LR, Hanser SF. 2002. Using information theory to assess the diversity, complexity, and development of communicative repertoires. Journal of Comparative Psychology. 116:166–172.

Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes. 4:792–794.

Mennill DJ, Battiston M, Wilson DR, Foote JR, Doucet SM. 2012. Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. Methods in Ecology and Evolution. 3:704–712.

Palmero AM, Espelosín J, Laiolo P, Illera JC. 2014. Information theory reveals that individual birds do not alter song complexity when varying song length. Animal Behaviour. 87:153–163.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics. 28:2537–2539.

Peet RK. 1975. Relative diversity indices. Ecology. 56:496–498.

Pielou EC. 1966. The measurement of diversity in different types of biological collections. Journal of Theoretical Biology. 13:131–144.

Pieretti N, Farina A, Morri D. 2011. A new methodology to infer the singing activity of an avian community: the acoustic complexity index (ACI). Ecological Indicators. 11:868–873.

Pijanowski BC, Villanueva–Rivera LJ, Dumyahn SL, Farina A, Krause BL, Napoletano BM, Gage SH, Pieretti N. 2011. Soundscape ecology: the science of sound in the landscape. BioScience. 61:203–216.

Ricklefs RE, Lau M. 1980. Bias and dispersion of overlap indices: results of some Monte Carlo simulations. Ecology. 61:1019–1024.

Sandoval L, Méndez C, Mennill DJ. 2014. Individual distinctiveness in the fine structural features and repertoire characteristics of the songs of white–eared ground–sparrows. Ethology. 120:275–286.

Shannon CE, Weaver W. 1949. The Mathematical Theory of Communication. Chicago: Il, University of Illinois.

Sherwin WB. 2010. Entropy and information approaches to genetic diversity and its expression: genomic geography. Entropy. 12:1765–1798.

Slabbekoorn H. 2004. Singing in the wild: the ecology of birdsong. In: Marler P, Slabbekoorn H, editors. Nature´s music, the science of bird song. San Diego: Elsevier Academic Press; p. 178–205.

Staicer CA, Spector DA, Horn AG. 1996. The dawn chorus and other diel patterns in acoustic signaling. In: Kroodsma DE, Miller EH, editors. Ecology and evolution of acoustic communication in birds. Ithaca: Comstock Publising Associates; p. 426–453.

Sueur J, Farina A. 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound. Biosemiotics 8:493–502.

Sueur J, Gasc A, Grandcolas P, Pavoine S. 2012. Global estimation of animal diversity using automatic acoustic sensors. In: Le Galliard JF, Guarini JM, Gaill F, editors. Sensors for ecology: towards integrated knowledge of ecosystems. Paris: CNRS; p. 99–117.

Truax B, Barrett GW. 2011. Soundscape in a context of acoustic and landscape ecology. Landscape Ecology. 26:1201–1207.

Ulanowicz RE. 1997. Ecology, the Ascendent Perspective. New York: NY, Columbia University Press.

Vargas–Castro LE, Sánchez NV, Barrantes G. 2012. Repertoire size and syllable sharing in the song of the clay–colored thrush (*Turdus grayi*). Wilson Journal of Ornithology. 124:446–453.

Wolda H. 1981. Similarity Indices, sample size and diversity. Oecologia. 50:296–302.

Table 1. Shannon-Weiner index of diversity (H') values for two populations with the same numbers of individuals, but different numbers of species.

Individuals	Species	H'	Number of individual per species
80	8	2	N1 = 10, N2 = 10, N3= 10, N4=10, N5 = 10, N6 = 10, N7 = 10, N8 = 10
80	16	2	N1 = 35, N2 = 6, N3 = 6, N4 = 5, N5 = 5, N6 = 5, N7 = 5, N8 = 5, N9 = 1, N10 = 1, N11 = 1, N12 = 1, N13 = 1, N14 = 1, N15 = 1, N16 = 1

Table 2. Results of Markov chain analyses for 20 individuals with eight song types in each individual’ repertoire. Shannon-Wiener entropy values (H') are also shown.

Individual	χ2	p	H'
i01	693.6	<0.001	2.08
i02	518.8	<0.001	1.88
i03	566.4	<0.001	1.95
i04	566.4	<0.001	1.95
i05	566.4	<0.001	1.95
i06	598.1	<0.001	1.92
i07	553.6	<0.001	1.48
i08	303.5	<0.001	1.55
i09	13.45	1	0.99
i10	0.57	1	0.39
i11	61.92	0.1	2.04
i12	58.16	0.27	2.06
i13	44.38	0.66	2.03
i14	67	0.04	2.06
i15	41.52	0.76	2
i16	51.29	0.38	2.06
i17	37.19	0.89	2.03
i18	41.7	0.76	2.02
i19	48.26	0.5	2.04
i20	39.28	0.83	2.05

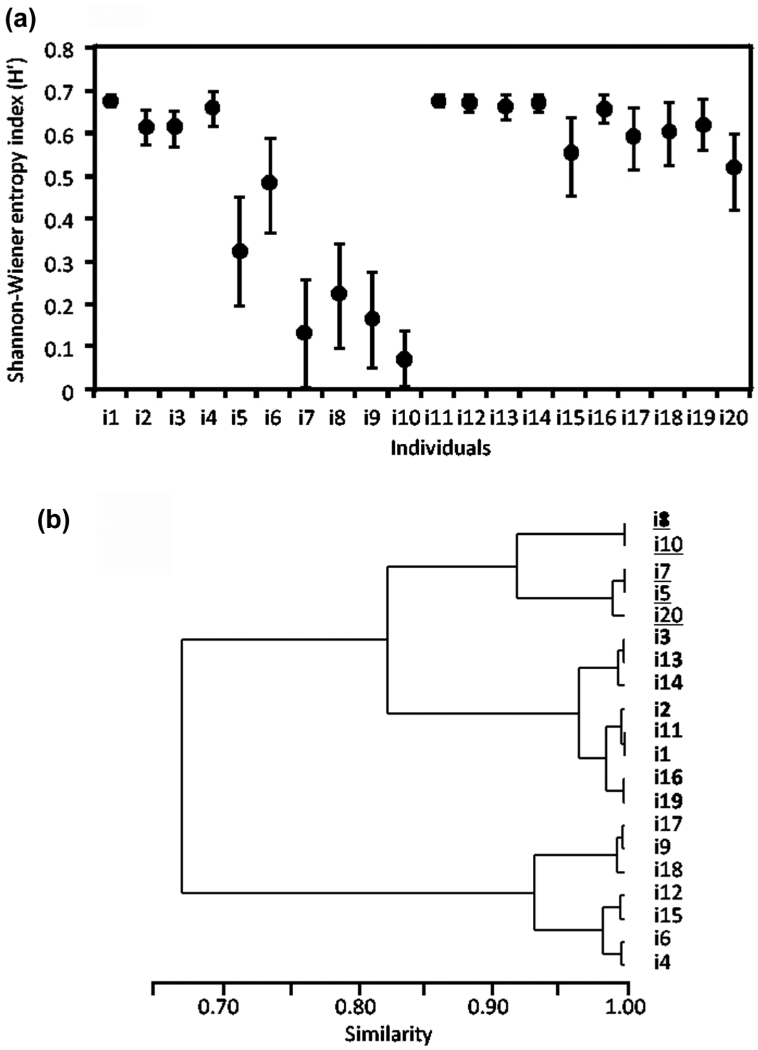

Figure 1. Analysis of 20 individuals with two song types in each individual's repertoire. (a) Results of the Shannon-Wiener entropy index. Error bars show 95% confidence intervals derived from bootstrapping. Individuals with overlapping error bars do not differ significantly in repertoire complexity. (b) Results of the Morisita similarity index. Groups used for the ANOSIM analysis (see methods) are denoted by different font type. Similarity is measured as the distance between the two individuals from their closest common node (represented by the similarity scale bar). Individuals separated only by a vertical line are identical to each other.

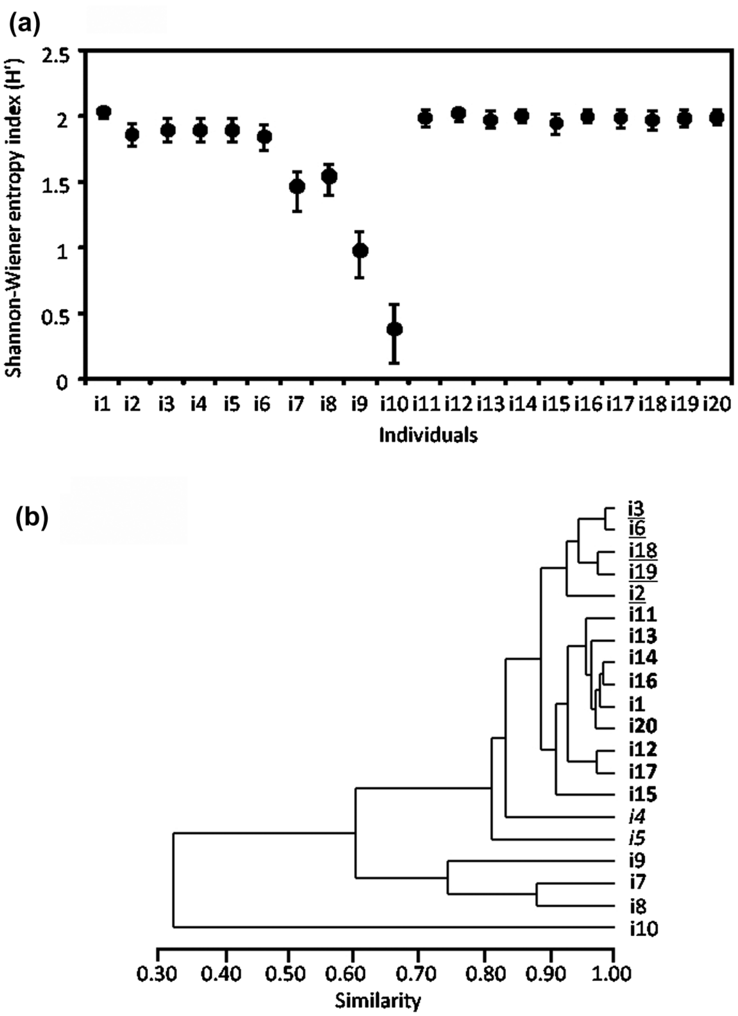

Figure 2. Analysis of 20 individuals with eight songs in each individual's repertoire. (a) Results of the Shannon-Wiener entropy index. Error bars show 95% confidence intervals derived from bootstrapping. Individuals with overlapping error bars do not differ significantly in repertoire complexity. (b) Results of the Morisita similarity index. Groups used for the ANOSIM analysis (see methods) are denoted by different font type. Similarity is measured as the distance between the two individuals from their closest common node (represented by the similarity scale bar). Individuals separated only by a vertical line are identical to each other.

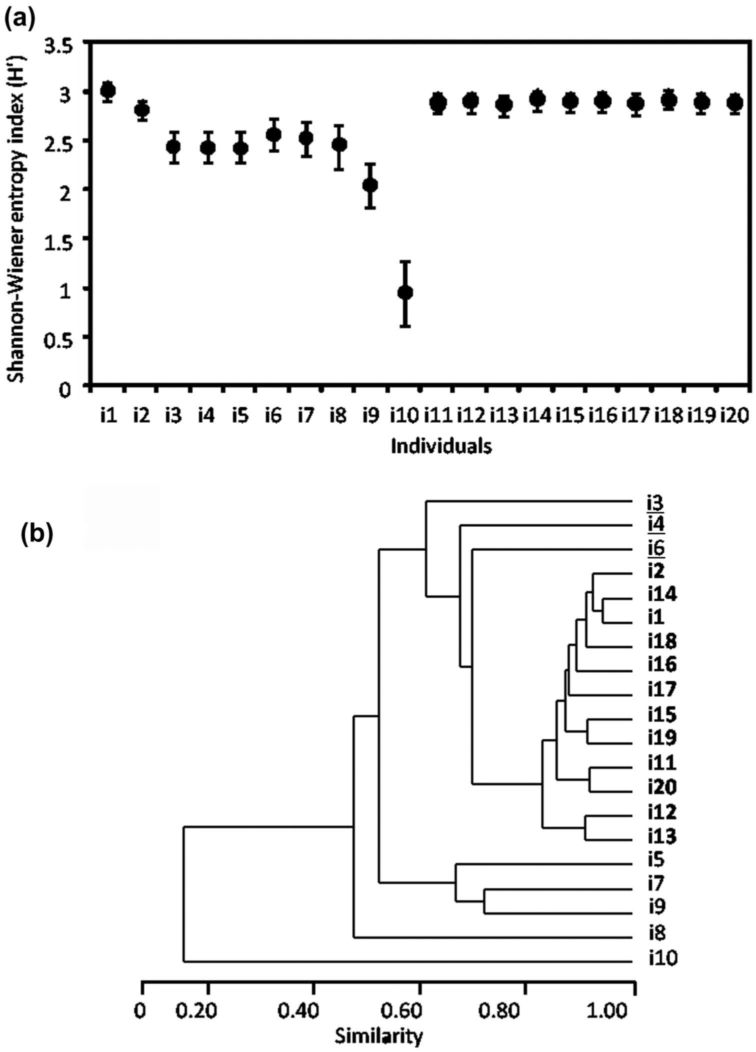

Figure 3. Analysis of 20 individuals with 20 songs in each individual's repertoire. (a) Results of the Shannon- Wiener entropy index. Error bars show 95% confidence intervals derived from bootstrapping. Individuals with overlapping error bars do not differ significantly in repertoire complexity. (b) Results of the Morisita similarity index. Groups used for the ANOSIM analysis (see methods) are denoted by different font type. Similarity is measured as the distance between the two individuals from their closest common node (represented by the similarity scale bar). Individuals separated only by a vertical line are identical to each other.

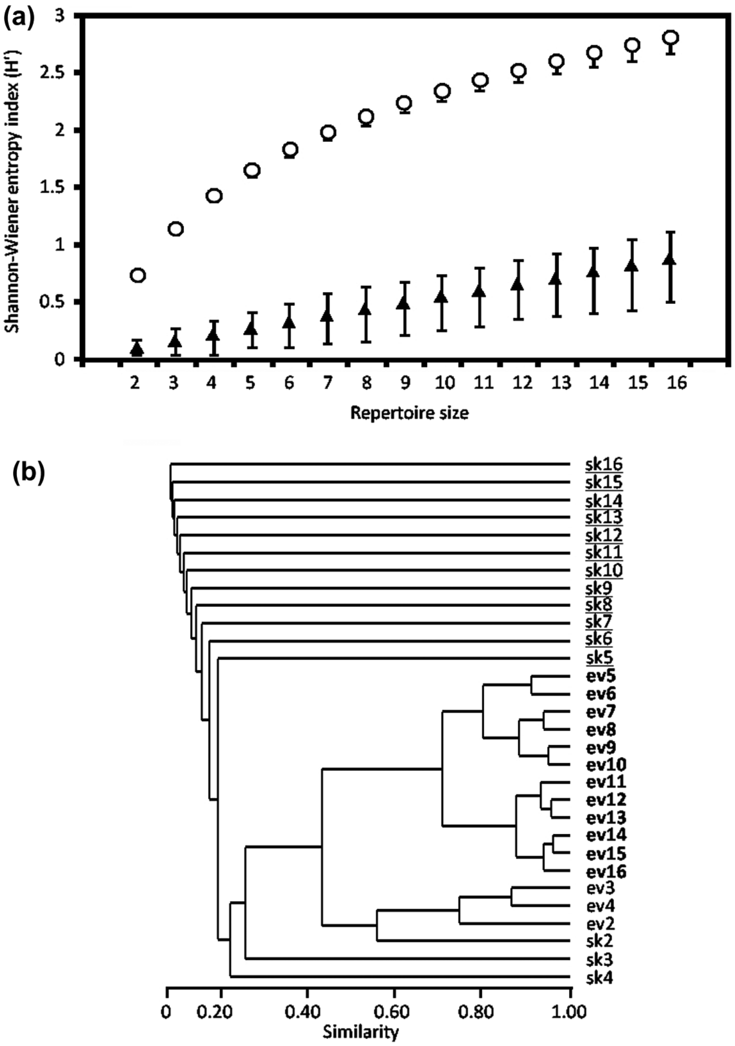

Figure 4. Analysis of 30 individuals with 2 to 16 song types in each individual's repertoire. (a) Results of the Shannon-Wiener entropy index. Error bars show 95% confidence intervals derived from bootstrapping. Individuals with overlapping error bars do not differ significantly in repertoire complexity. (b) Results of the Morisita similarity index. Groups used for the ANOSIM analysis (see methods) are denoted by different fonts. Similarity is measured as the distance between the two individuals from their closest common node (represented by the similarity scale bar). Individuals separated only by a vertical line are identical to each other. Individuals whose songs were distributed evenly among song types are represented by circles (panel a) or the prefix 'ev' (panel b), whereas individuals whose songs were distributed among song types according to a skewed distribution are represented by triangles (panel a) or the prefix 'sk' (panels b, c).

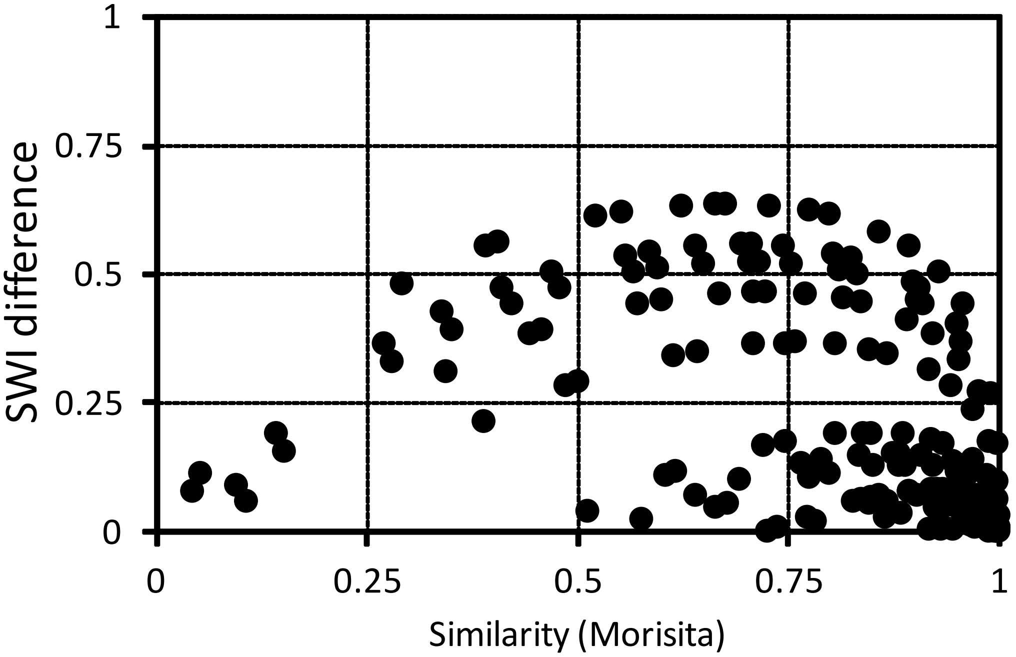

Figure S1. Relationship between Morisita index of similarity and the differences in the Shannon-Wiener entropy index values for 20 individuals with two song types in each individual's repertoire. Dots represent all pairwise comparisons. Morisita index values near 1 indicate that repertoires are similar, whereas index values near 0 indicate that repertoires are different.

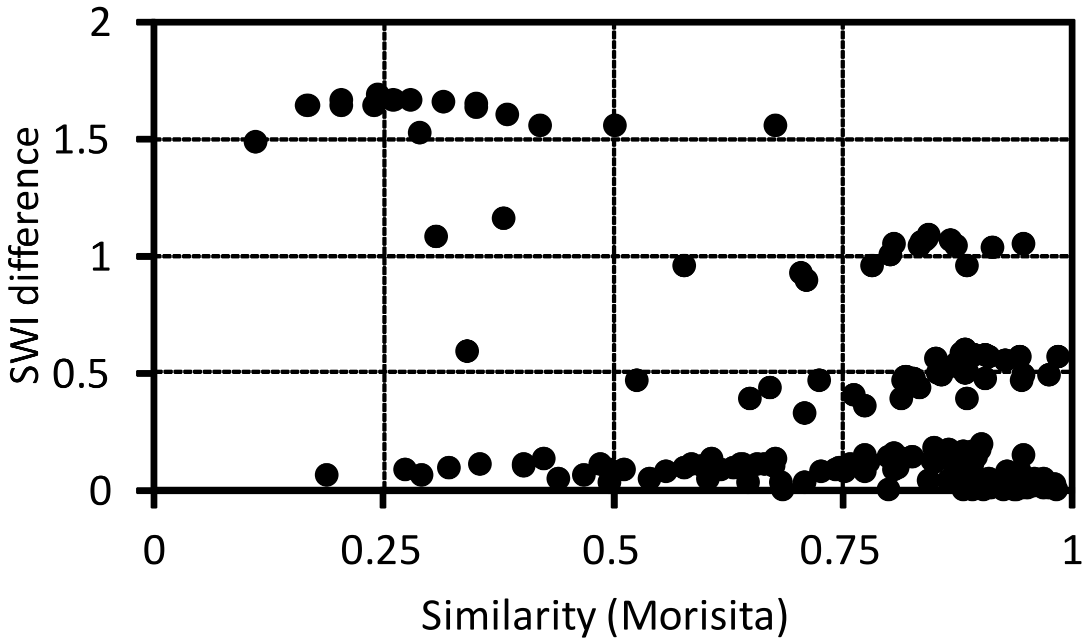

Figure S2. Relationship between Morisita index of similarity and the differences in the Shannon-Wiener entropy index values for 20 individuals with eight song types in each individual's repertoire. Dots represent all pairwise comparisons. Morisita index values near 1 indicate that repertoires are similar, whereas index values near 0 indicate that repertoires are different.

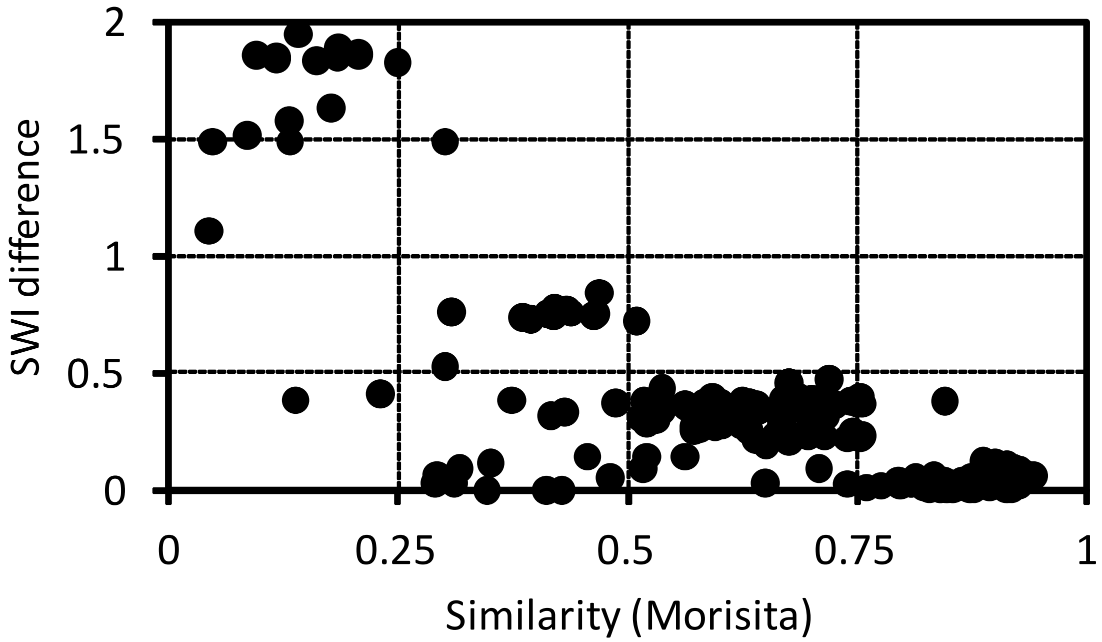

Figure S3. Relationship between Morisita index of similarity and the differences in the Shannon-Wiener entropy index values for 20 individuals with 20 song types in each individual's repertoire. Dots represent all pairwise comparisons. Morisita index values near 1 indicate that repertoires are similar, whereas index values near 0 indicate that repertoires are different.

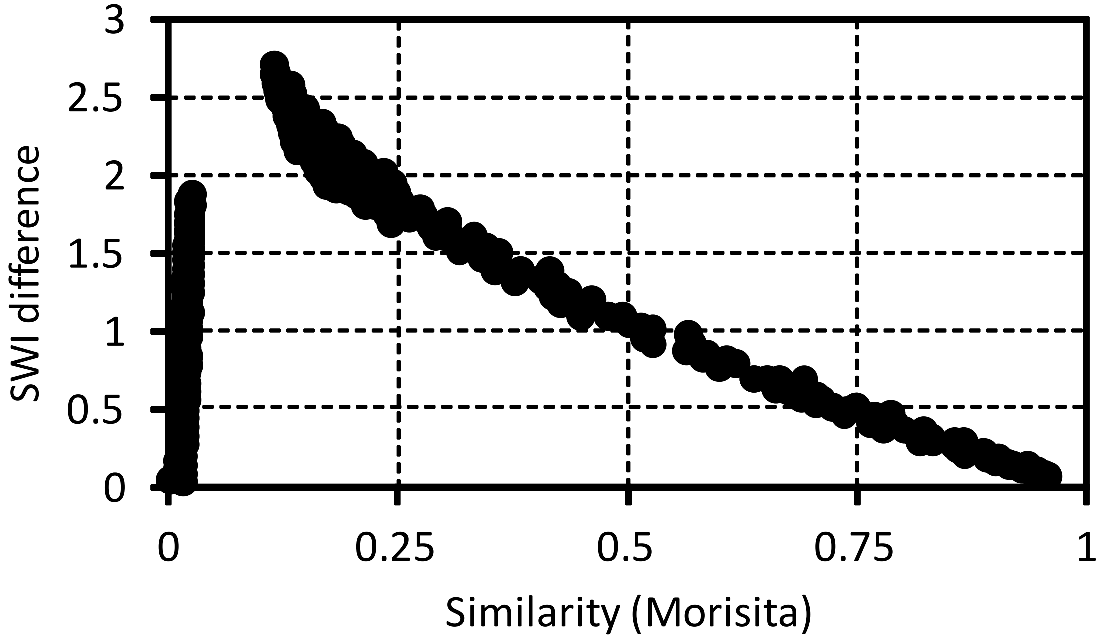

Figure S4. Relationship between Morisita index of similarity and the differences in the Shannon-Wiener entropy index values for 30 individuals with 2 to 16 song types in each individual's repertoire. Dots represent all pairwise comparisons. Morisita index values near 1 indicate that repertoires are similar, whereas index values near 0 indicate that repertoires are different.

Table S1. Data used under the first scenario of repertoire complexity, where each individual has two song types in its repertoire. The individuals in each of the three groups used for the ANOSIM analysis (see methods) are: group 1 (i5, i7, i8, i10 and i20), group 2 (i1, i2, i3, i11, i13, i14, i16, and i19), and group 3 (i4, i6, i9, i12, i15, i17, and i18).

	Individuals																			
Sound type	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
a	50	1	30	70	40	10	20	97	6	96	51	54	42	46	25	59	69	34	61	77
b	50	99	70	30	60	90	80	3	94	4	49	46	58	54	75	41	31	66	39	23
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table S2. Data used under the second scenario of repertoire complexity, where each individual has eight song types in its repertoire. The individuals in each of the three groups used for the ANOSIM analysis (see methods) are: group 1 (i2, i3, i6, i18, and i19), group 2 (i1, i11, i12, i13, i14, i15, i16, i17, and i20), and group 3 (i4, i5, i7, i8, i9, and i10).

	Individuals																			
Sound type	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
a	13	1	5	15	20	1	2	1	3	2	16	12	9	9	7	11	8	9	9	9
b	13	1	5	15	20	1	2	1	3	2	10	13	14	13	15	12	10	14	8	12
c	13	1	5	15	5	1	2	1	4	16	14	13	14	16	16	12	15	5	9	17
d	13	1	15	20	5	1	2	1	18	16	14	7	18	16	17	17	7	16	14	13
e	12	1	15	20	5	1	2	24	18	16	9	15	16	11	12	9	13	10	13	13
f	12	1	15	5	15	1	30	24	18	16	19	12	14	15	7	16	17	13	13	13
g	12	1	20	5	15	47	30	24	18	16	8	15	7	10	6	9	18	15	16	7
h	12	93	20	5	15	47	30	24	18	16	10	13	8	10	20	14	12	18	18	16
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table S3. Data used under the third scenario of repertoire complexity, where each individual has twenty song types in its repertoire. The individuals in each of the three groups used for the ANOSIM analysis (see methods) are: group 1 (i3, i4, and i6), group 2 (i5, i7, i8, i9, and i10), and group 3 (i1, i2, i11, i12, i13, i14, i15, i16, i17, i18, i19 and i20).

	Individuals																			
Sound type	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
a	5	1	5	1	13	26	2	3	3	20	4	7	10	2	6	6	5	2	3	5
b	5	1	5	1	5	7	5	2	3	20	7	6	5	5	7	5	4	3	8	5
c	5	1	5	1	13	9	2	3	3	20	3	2	4	5	5	3	4	9	5	1
d	5	1	5	1	5	3	5	4	3	20	4	4	5	6	3	4	4	5	5	5
e	5	1	5	1	13	2	2	3	3	2	8	4	6	6	5	3	2	5	5	7
f	5	1	1	1	1	5	5	4	3	2	4	7	3	4	1	4	6	3	3	6
g	5	1	1	1	13	6	2	4	3	2	9	4	3	4	4	7	4	4	6	10
h	5	1	1	1	5	1	5	6	3	2	4	8	7	4	7	2	7	4	3	2
i	5	1	1	1	5	2	10	4	3	1	4	6	7	4	6	2	4	8	6	3
j	5	1	1	1	1	7	15	2	3	1	7	7	5	3	3	6	1	3	2	5
k	5	1	1	13	13	4	18	3	7	1	7	5	3	4	9	8	9	8	11	6
l	5	1	1	13	1	2	3	7	7	1	9	2	2	8	6	6	4	6	2	6
m	5	1	1	13	1	1	2	33	7	1	2	7	8	4	9	8	2	5	6	6
n	5	1	1	13	1	3	1	3	7	1	5	5	3	6	5	4	5	3	7	7
o	5	1	1	13	1	2	1	4	7	1	3	4	3	3	6	6	8	5	5	5
p	5	1	13	5	5	2	4	2	7	1	6	2	2	7	3	5	10	7	2	8
q	5	1	13	5	1	2	3	2	7	1	2	2	8	8	3	8	4	4	3	4
r	5	1	13	5	1	5	2	3	7	1	2	4	4	7	6	3	8	4	5	4
s	5	1	13	5	1	6	6	5	7	1	6	6	2	3	3	2	3	5	6	1
t	5	81	13	5	1	5	7	3	7	1	4	8	10	7	3	8	6	7	7	4
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table S4. Data used under the fourth scenario of repertoire coomplexity. * ev = even distribution of the sound types. sk = skewed distribution of the sound types.

	Individuals*																													
Sound type	ev2	sk2	ev3	sk3	ev4	sk4	ev5	sk5	ev6	sk6	ev7	sk7	ev8	sk8	ev9	sk9	ev10	sk10	ev11	sk11	ev12	sk12	ev13	sk13	ev14	sk14	ev15	sk15	ev16	sk16
a	50	1	33	1	25	1	20	1	16	1	14	1	13	1	11	1	10	1	9	1	8	1	7	1	7	1	6	1	6	1
b	50	99	33	1	25	1	20	1	16	1	14	1	13	1	11	1	10	1	9	1	8	1	7	1	7	1	6	1	6	1
c	0	0	34	98	25	1	20	1	17	1	14	1	13	1	11	1	10	1	9	1	8	1	7	1	7	1	6	1	6	1
d	0	0	0	0	25	97	20	1	17	1	14	1	13	1	11	1	10	1	9	1	8	1	7	1	7	1	6	1	6	1
e	0	0	0	0	0	0	20	96	17	1	14	1	12	1	11	1	10	1	9	1	8	1	8	1	7	1	6	1	6	1
f	0	0	0	0	0	0	0	0	17	95	15	1	12	1	11	1	10	1	9	1	8	1	8	1	7	1	7	1	6	1
g	0	0	0	0	0	0	0	0	0	0	15	94	12	1	11	1	10	1	9	1	8	1	8	1	7	1	7	1	6	1
h	0	0	0	0	0	0	0	0	0	0	0	0	12	93	11	1	10	1	9	1	8	1	8	1	7	1	7	1	6	1
i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	92	10	1	9	1	9	1	8	1	7	1	7	1	6	1
j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	91	9	1	9	1	8	1	7	1	7	1	6	1
k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	90	9	1	8	1	7	1	7	1	6	1
l	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	89	8	0	7	1	7	1	6	1
m	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	88	8	1	7	1	7	1
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	87	7	1	7	1
o	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	86	7	1
p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	85
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99	100	100	100	100	100	100