TITLE:

Measuring acoustic complexity in continuously varying signals: how complex is a wolf howl?

AUTHORS

Arik Kershenbaum[1], Éloïse C. Déaux[2], Bilal Habib[3], Brian Mitchell[4], Vicente Palacios[5], Holly Root-Gutteridge[6], Sara Waller[7]

[1]Department of Zoology, University of Cambridge, UK

[2] Department of Biological Sciences, Macquarie University, Sydney, 2109, Australia

[3] Department of Animal Ecology and Conservation Biology, Wildlife Institute of India, Dehradun, India

[4] The Rubenstein School of Environment and Natural Resources, University of Vermont, USA

[5]Instituto Cavanilles de Biodiversidad y Biología Evolutiva. University of Valencia, Spain

[6]Department of Biology, Syracuse University, USA

[7]Montana State University, Bozeman, MT, USA

CORRESPONDING AUTHOR: Arik Kershenbaum, arik.kershenbaum@gmail.com

ABSTRACT

Communicative complexity is a key behavioural and ecological indicator in the study of animal cognition. Much attention has been given to measures such as repertoire size and syntactic structure in both bird and mammal vocalisations, as large repertoires and complex call combinations may give an indication of the cognitive abilities both of the sender and receiver. However, many animals communicate using a continuous vocal signal that does not easily lend itself to be described by concepts such as "repertoire". For example, dolphin whistles and wolf howls both have complex patterns of frequency modulation, so that no two howls or whistles are quite the same. Is there a sense in which some of these vocalisations may be more "complex" than others? Can we arrive at a quantitative metric for complexity in a continuously varying signal? Such a metric would allow us to extend familiar analyses of communicative complexity to those species where vocal behaviour is not restricted to sequences of stereotyped syllables. We present four measures of complexity in continuous signals (Wiener Entropy, Autocorrelation, Inflection Point Count, and Parsons Entropy), and examine their relevance using example data from members of the genus *Canis*. We show that each metric can lead to different conclusions regarding which howls could be considered complex or not. Ultimately, complexity is poorly defined and researchers must compare metrics to ensure that they reflect the properties for which the hypothesis is being tested.

KEYWORDS: Autocorrelation, Canids, Communication, Complexity, Entropy

INTRODUCTION

Vocal complexity is considered an important property of animal communication (Freeberg and Krams 2015; Larson 2004; McCowan et al. 2002; Pollard and Blumstein 2012), despite being poorly defined, with little agreement how complexity should be quantified (Edmonds 1999). Despite this, complexity has been used to explain different aspects of animal behaviour. For instance, evidence exists in several bird species that females choose mates at least partially on the basis of the complexity of male song (Darolová et al. 2012; Hiebert et al. 1989), and other males may use complexity cues to make conflict escalation decisions (Leitão et al. 2006). It has been postulated that birdsong complexity acts as an index signal; being positively correlated with nutritional competence and cognitive abilities, and negatively correlated with early life stress (Nowicki et al. 2002). Similarly, recent studies have suggested that a correlation exists between communicative complexity and social complexity, such that species with more complex social systems also have more complex communicative interactions (Freeberg and Krams 2015; Krams et al. 2012; Pollard and Blumstein 2012). This in turn could shed light on possible evolutionary pathways to the development of language as an adaptation of highly complex social groups in early hominins (Seyfarth and Cheney 2014).

Communicative complexity can also have practical implications for the identification of and discrimination between similar sub-populations where complexity varies geographically (Briefer et al. 2010; Kershenbaum et al. 2012). While many species can distinguish between the vocalisations of in-group and out-group individuals, e.g. wolves *Canis lupus* (Palacios et al. 2015; Zaccaroni et al. 2012), elephants *Loxodonta africana* (O'Connell-Rodwell et al. 2007), and multiple bird species (Nakagawa et al. 2001; Radford 2005), it is often not clear what vocal cues are being used to make this discrimination, and complexity characteristics may play a role (Briefer et al. 2008). Geographic differences in vocal complexity may be particularly noticeable where ecological conditions lead to differences in food availability, cognitive development, and hence vocal repertoire size (Byers and Kroodsma 2009; Kipper et al. 2006; Pfaff et al. 2007). In parallel, researchers can make use of differences in repertoire size, for example, to distinguish between sub-populations of birds and mammals (Gwilliam et al. 2008; Pitcher et al. 2012). Clearly, vocal complexity is an important phenomenon with far ranging implications for the study of animal communication.

Previous studies of vocal complexity have focussed largely on birdsong, because of three essential properties that make this communication modality particularly tractable: (a) most birdsong can be divided into discrete syllables or notes (Marler and Slabbekoorn 2004); (b) there exists a simple metric - repertoire size - for measuring purported complexity (Byers and Kroodsma 2009); and (c) the well-established role of birdsong in mate choice provides the opportunity for manipulative as well as correlative experiments to be carried out, quite clearly demonstrating the role of song complexity in enhancing fitness (Searcy 1992). Even when birdsong is open-ended so that repertoire cannot adequately be defined, for example in the northern mockingbird *Mimus polyglottos* (Gammon 2014), the discrete nature of the song syllables means that other measurements of communicative complexity can be used, most notably Shannon entropy (Briefer et al. 2010; Da Silva et al. 2000; Kershenbaum 2013). Although some animals from other taxa have vocal communication systems that are similarly discrete, e.g. rock hyrax *Procavia capensis* (Kershenbaum et al. 2012), or are closed-ended, e.g. several primates (Cäsar et al. 2012; Zuberbühler 2002), outside of passerine birds they are the exception rather than the rule. Indeed, the vocal communication system of some species consists entirely of signals whose properties are continuously varying, and with such signals the existing concepts of complexity (e.g. based on repertoire and entropy) cannot be applied. For example, considerable empirical evidence supports the existence of semantic information in the whistles of bottlenose dolphins *Tursiops truncatus*; in particular, the use of signature whistles to signal individual identity (Fripp et al. 2005; Kershenbaum et al. 2013; Quick and Janik 2012; Sayigh et al. 1999). However, dolphin whistles are relatively long and unbroken tonal signals that are continuously frequency modulated. Similarly, wolf howls are continuous frequency modulated signals that have been shown to contain individual identity information (Palacios et al. 2007; Root-Gutteridge et al. 2014; Tooze et al. 1990), but cannot be classified into stereotyped categories. One possible interpretation of complexity, which may or may not be intuitive to the reader, is that the frequency of a *simple* signal varies little, or predictably, with time, whereas a *complex* signal varies greatly, and unpredictably. The reader may examine Figure 1, which shows several examples of wolf howls and decide which howls are simple and which complex. However, although some howls may appear intuitively more complex than others, the lack of an objective definition of complexity for such signals renders the judgement unhelpful. As such, previous studies of information content in continuously varying animal vocal signals have had to make use of alternative techniques, example for measuring similarity between pairs of vocalisations, rather than quantifying characteristics of the vocalisations themselves (Kershenbaum et al. 2013).

Complexity in itself is poorly defined (Edmonds 1999), and as a result any particular use of the term in examining animal behaviour is liable to be criticised. In particular, information theoretical definitions of complexity based on concepts such as entropy are often regarded with suspicion by biologists, because the most complex (highest entropy) signals are in fact random signals - something that most ethologists would consider to be non-complex (Suzuki et al. 2005). We agree that defining complexity is difficult, however we hope to mitigate this difficulty somewhat by insisting that researchers should examine and understand what kind of signal a particular definition of complexity would deem either complex or non-complex. It would then be possible to determine whether a particular definition of complexity meets the demands of discriminating between signals in relation to the study's hypotheses.

In this paper, we aim to show how concepts of continuous complexity can be measured using different approaches, and we illustrate how use of a particular complexity metric can lead to different conclusions from using other metrics. We present four candidate complexity metrics and compare their performance against each other, identifying which kinds of signals each particular metric would indicate to be "complex" or "not complex". We recognise that in drawing up a proposed metric of complexity for continuous signals, it is inevitable that subjective interpretations of the complexity or simplicity of a signal must necessarily influence the decisions a researcher makes in designing an experiment. Rather than attempt to avoid this subjective tendency, we hope to formalise it somewhat, by presenting a range of complexity metrics, along with illustration of their significance for signals of different types. That way, researchers can choose a metric that captures the "kind of" complexity for which they are searching; providing a quantitative measure for an essentially subjective property.

METHODS

We describe below four quantitative metrics that have been previously used for quantifying complexity in continuous signals: Wiener Entropy, Autocorrelation, Inflection Point Count, and Parsons Entropy. For each of these metrics, we define how they are computed and explain in what way they could be considered to be measures of complexity in a continuous signal, giving detailed examples using simulated waveforms as shown in Figure 2. As each of our proposed metrics measures quite different features of acoustic signals, it is constructive to illustrate the behaviour of each metric by showing examples of signals with both high and low metric values, as this provides an indication of which signal features are being emphasised by each metric. We do this using example acoustic signals drawn from an empirical data set consisting of 2,014 coded canid howls from 16 different species and subspecies, as described in a previous work (Kershenbaum et al. 2016). We calculated each complexity metric for each howl and then plot on a time-frequency graph the three howls with the lowest metric values, and the three howls with the highest values.

Wiener Entropy

Wiener Entropy, or spectral flatness, has been proposed as a measure of the complexity of birdsong elements, and has been applied to the analysis of increasing complexity during the process of song learning in juvenile birds (Baker and Logue 2003; Tchernichovski et al. 2000). Wiener Entropy is a measure of the extent to which a signal contains a mixture of frequencies, as opposed to a single frequency or tone. As such, a value of 0 represents a single sine wave, and a value of 1 indicates white noise, in which all frequencies are equally represented. To calculate the Wiener Entropy, we first perform a Fourier transform to calculate the spectral power P present at each of N distinct frequency bins. The formal definition is given as the ratio of the geometric means of the spectral powers to the arithmetic mean:

$$WE = \frac{\sqrt[N]{\prod_{f=1}^{N} P(f)}}{\frac{1}{N}\sum_{f=1}^{N} P(f)}$$

Wiener Entropy can be applied to the signal spectrum, where *P(f)* corresponds to the FFT of the input waveform (column C in Figure 2), or to the spectrographic representation of the signal, where *P(f)* corresponds to *F(t)* (column B in Figure 2). The former definition also measures the entropic contribution of background noise and discards any information on temporal variation in frequency, therefore, we calculate Wiener Entropy only on the dominant signal frequency, i.e. column B in Figure 2. In addition, we square-root transform this metric for normality.

Autocorrelation

Autocorrelation (Figure 2, column D) measures the self-similarity of a signal, and so quantifies the extent to which the signal contains repetitions of the same pattern, or is varied without similarity (Stoica and Moses 2005). The autocorrelation sequence of a signal consisting of *N* discrete samples, *F(1...N)* is measured by shifting the signal by time lag *l* and calculating the correlation between the shifted and the unshifted signals:

$$ac(l) = \sum_{n=0}^{N} F(n) \cdot F(n-l)$$

The Autocorrelation metric is then calculated as the sum of *ac(l)* for all *l*. As with Weiner entropy, we measure the repetitiveness of the howl frequency modulation, rather than flatness *per se*, by calculating autocorrelation on the dominant signal frequency in the spectrogram. In addition, we log-transform this metric for normality.

Inflection points

A number of studies, particularly with cetaceans, have measured vocal complexity by counting the number of inflection points in a vocalisation (Janik et al. 1994; May-Collado and Wartzok 2008). A more complex signal, in this context, is a signal in which the frequency is changing direction (rising/falling) often (e.g. Figure 2, panel 4B). In keeping with these studies, we define the number of inflection points as the count of changes in gradient direction of the vocalisation. To ensure that this complexity metric takes continuous values, we divide the number of inflection points by the length of the signal, which also standardises the metric, to remove the correlation between vocalisation length and complexity that would otherwise be present. In addition, we square-root transformed this metric for normality. Simplified Matlab code for counting the number of inflection points in a vector **X** is shown below:

```
1. g=gradient(X);
2. s=sign(g);
3. d=g(1:end-1)~=g(2:end);
4. C=sum(d);
```

The above algorithm (1) measures the gradient at each point, (2) determines the sign of the gradient (positive, increasing, or negative, decreasing), (3) tests whether the sign of the gradient at this point is different from the sign of the gradient at the next point, which would indicate a change in direction, and (4) counts the number of such changes in direction.

Parsons Entropy

The Parsons code is a reduced representation of a varying signal, used primarily for music retrieval systems (Downie 2003; Parsons and Levin 1975). However, a recent study showed that dolphin signature whistles can be represented as Parsons codes, while maintaining much of the individual identity information (Kershenbaum et al. 2013). To convert a continuous signal to a Parsons code, we divide the signal into a fixed number of slices, and record whether the frequency from one slice to the next is rising, falling, or remaining constant. To increase the descriptive power of the Parsons code, we can distinguish between rises and falls of different magnitude, classifying the 10% of the largest magnitude changes as “big rise” or “big drop” (Müllensiefen and Frieler 2004; Pauws 2002), and thus dividing changes in pitch into seven categories: big rise, medium rise, small rise, no change, small drop, medium drop, big drop (Figure 3). The implementation of this coding is described more fully in (Kershenbaum et al. 2013). The continuous signal has now been converted into a series of discrete characters from a finite alphabet of seven step categories, and so we can then calculate the simple

Shannon entropy (Cover and Thomas 1991) as is often done with stereotyped signals, where *P(n)* is the probability of occurrence of step category *n*. We refer to this entropy metric that measures the unpredictability of the Parsons code as the "Parsons Entropy".

$$PE = -\sum_{n=1}^{7} P(n) \log_7 P(n)$$

As each of the four metrics Wiener Entropy, Autocorrelation, Inflection Point Count, and Parsons Entropy are all purported to measure the same property - signal complexity - the metrics may potentially measure similar features of the acoustic signals. Therefore, we test directly for correlation between the different metrics by calculating the Pearson's correlation coefficient between each pair of metrics. All analyses were carried out in Matlab R2014b (The Mathworks, Inc).

RESULTS

Complexity of simulated waveforms

The values of the four metrics for each of the five waveforms in Figure 2 are shown in Figure 4. All metrics gave the constant frequency (1) the lowest score, representing the simplest waveform. The random waveform (5) received the highest scores from three of the metrics, Wiener Entropy, Inflection Point Count, and Parsons Entropy, indicating (as with traditional entropy measures) that randomness is interpreted as high complexity. Autocorrelation was low for the random waveform. The oscillating frequency (4) scored highly for complexity with all metrics (especially Autocorrelation and Parsons Entropy), consistent with the intuitive interpretation of this as a complex signal. However, the two-frequency waveform (2) received similar Wiener Entropy and Autocorrelation scores to constantly increasing frequency (3), but notably higher scores than (3) for Inflection Point Count and Parsons Entropy. Indeed, both Inflection Point Count and Parsons Entropy considered the constantly increasing frequency (3) to be approximately as complex as the constant frequency (1).

Complexity of wolf howls

Examples of howls with the lowest, and highest values of each of the complexity metrics are shown in Figure 5. Wiener Entropy and Autocorrelation illustrate the difficulty of traditional metrics as indicators of complexity. It is not clear that the howls with the lowest Wiener Entropy or Autocorrelation are necessarily less complex by subjective interpretation than those with high Wiener Entropy or Autocorrelation. A low Wiener Entropy score is achieved by a signal possessing a single frequency (i.e. approximating a part of a sine wave), whereas high Wiener Entropy score is achieved by flat signals, which are transformed by FFT to a mixture of a large number of frequencies, and hence high entropy. Neither appear to be particularly complex by intuitive definition. Low Autocorrelation scores are achieved by irregular but not repetitive signals, and such irregularity is a promising trait of complexity; however high Autocorrelation scores are achieved by signals with a single frequency, which does not appear to be either complex or simple.

In contrast, the number of inflection points seems an intuitive measure of complexity, as high Inflection Point Count howls are very varied, whereas low Inflection Point Count howls appear simpler. However, some howls with low Inflection Point Count still have considerable variation in frequency. Parsons Entropy also detects subjectively complex howls, and those with low Parsons Entropy appear subjectively simple; specifically, the frequency of these simple howls varies monotonically.

Colinearity among complexity metrics

Weak correlations existed between all metrics (Figure 6), however, a stronger negative correlation was found between Wiener Entropy and Parsons Entropy (R=-0.38). Taken with the other differences found between metrics, there does not appear to be grounds for describing any pair of metrics as co-varying, and the metrics appear to be measuring different aspects of complexity.

DISCUSSION

We have illustrated the similarities and differences between four different metrics, each of which could be considered a quantitative measure of complexity in a continuously varying signal. Despite the poorly defined nature of signal complexity, we have provided the reader with both quantitative comparisons, and qualitative illustrations of the result of using each of these metrics in the evaluation of the complexity of simulated signals and natural of canid howls.

All metrics distinguished clearly between a constant flat frequency, and a randomly varying signal, with all metrics except Autocorrelation placing the random signal at the most complex end of the quantitative scale. Researchers will need to consider whether or not the characterisation of a random signal as "complex" (a definition taken from the field of entropy and information theory) is consistent with the hypotheses that they are testing. In contrast to Wiener Entropy, Inflection Point Count, and Parsons Entropy, Autocorrelation gave a higher score to a regularly varying signal than to the random one. Parsons Entropy also gave a much higher value for a regularly varying signal than for the flat, rising, and random frequencies, indicating that Parsons Entropy, which measures changes in slope, is a good metric for measuring the extent to which a signal changes with time - either regularly or irregularly. When examining actual howls qualitatively, both Inflection Point Count and Parsons Entropy appeared to distinguish between howls that the authors felt looked "simple" (i.e. varying in frequency in a constant way) and those that looked "complex" (i.e. varying in an inconsistent way with time), although clearly this subjective distinction may not be globally applicable. When examining the example howls for low and high Wiener Entropy and Autocorrelation, there did not appear to be as much of a subjective difference in complexity. However, the essence of these results is to provide the comparison, rather than to impose subjective conclusions, and Figure 5 makes this comparison clearly.

Despite the fact that complexity in vocal signals of any kind remains poorly defined (Kershenbaum 2013), the concept of complexity is still widely used for investigating questions of proximal behaviour (Darolová et al. 2012; Demartsev et al. 2014; Gustison and Bergman 2016), ultimate fitness (Freeberg et al. 2012; Ord et al. 2012; Pollard and Blumstein 2012), and drivers of the evolution of social systems (Bergman and Beehner 2015; Freeberg and Krams 2015; Krams et al. 2012). Even if we accept the definition of entropy as a proxy for complexity (Doyle et al. 2008), it is not clear how such a metric can be applied to continuously varying signals. We have shown that multiple approaches are possible, each with its benefits and disadvantages. For example, counting the number of inflection points is a useful method for identifying highly frequency modulated signals, but can become overwhelmed in the presence of a highly random signal. Parsons Entropy may suffer less from this constraint, as the signal is divided into discrete segments. Wiener Entropy measures the noisiness of a signal, but can misinterpret a simple upsweep as complex as it contains many frequencies, albeit spread through time. Autocorrelation is a powerful tool for detecting repetition, but returns a low value for asymmetric changes in frequency.

We have no objective gold-standard of complexity to compare to our metrics and to indicate whether a particular metric truly captures the property of complexity or not. Yet we believe that our study has merit precisely because it allows quantification of the subjective measure for which researchers may be searching. If an oscillating signal is the nature of complexity being tested, then Inflection Point Count or Parsons Entropy may be the best metric to use. However, if for a particular hypothesis, randomness is best rejected as not complex, then perhaps Autocorrelation is a better-suited metric. It is vital for researchers to understand the implications of their choice of a particular complexity metric, rather than to make use of a metric whose properties may be unknown, and perhaps surprising. Our approach of defining a quantitative metric also has the advantage of enabling clearer comparisons between different study systems that may use similar but non-quantitative assessments of vocalisation type. Qualitative descriptions of frequency modulations in continuous signals, e.g. "flat", "rising", "step up" (Hallberg 2007; Palacios et al. 2007) are useful, but difficult to compare between studies.

Vocal signal complexity is likely an important property in the communication of many species, including birds (Briefer et al. 2010; Darolová et al. 2012; Freeberg 2008; Kipper et al. 2006; Leitão et al. 2006), amphibians (Larson 2004; Narins and Capranica 1978), terrestrial mammals (Demartsev et al. 2014; Gustison and Bergman 2016; Schlenker et al. in press) and cetaceans (Doyle et al. 2008; Ferrer-i-Cancho and McCowan 2009; Garland et al. 2013; Nash and Bowles 2011), and also human language (Ferrer-i-Cancho 2005; Ferrer-i-Cancho and Solé 2003; Kershenbaum et al. 2014; Montemurro and Zanette 2015). Complexity is a difficult property to measure, particularly when the signals are continuously varying, rather than a sequence of discrete notes. To investigate questions such as the connection between social complexity and vocal complexity, appropriate measures of complexity must be found (Freeberg and Krams 2015; Kershenbaum 2013). Species with continuously varying vocalisations, such as wolves and dolphins, share many of the properties of particular interest in the investigation of the evolution of communication. Wolves and dolphins in particular are highly social and cooperative, as well as having intelligent problem-solving abilities (wolves: Mech and Boitani 2010; Mech et al. 2015, dolphins: Gazda et al. 2005; King and Janik 2015; Wells 2003), and so it would be unfortunate if a quantitative assessment of their vocal complexity were neglected.

We have shown that multiple metrics do exist for capturing the complexity of these vocalisations, and we have provided quantitative tools to assess the suitability of the different types of metrics. We encourage researchers to make use of such quantitative measures when testing hypotheses in these and similar species, thereby extending the investigation of complex communication in animals to those species not previously amenable to quantitative analysis.

ACKNOWLEDGEMENTS

AK thanks Sheela Hira and Knoxville Zoo, Laura Pearson and Asheville Zoo, and Chris Lasher and North Carolina Zoo for access to record their red wolves, and Arthur Povey for assistance coding the data. Recording work was approved by the Institutional Animal Care and Use Committee of the University of Tennessee. AK is supported by a Herchel Smith postdoctoral fellowship at the University of Cambridge. Part of this work was carried out while AK was a Postdoctoral Fellow at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. BH is thankful to the State Forest Departments of Himachal Pradesh, J&K, and Maharashtra, and to various zoos in India for permitting us to record howls. HRG is grateful to all who helped with the project: the staff at Colchester Zoo; the Wildwood Trust, the Borror Laboratory of Bioacoustics; the British Library; Lupus Laetus; Polish Mammal Research Institute; Tigress Productions; the BBC Natural History Unit; Longleat Safari Park; Tierstimmen Archiv; Wild Sweden; Wolf Park; the Macaulay Sound Library and the UK Wolf Conservation Trust; and Mike Collins, Teresa Palmer, Monty Sloan, Karl-Heinz Frommolt, Yorgos Iliopoulos, Christine Anhalt, Louise Gentle, Richard Yarnell, Victoria Allison Hughes and Susan Parks. BRM thanks the USDA/APHIS/WS/National Wildlife Research Center for supporting his doctoral research and providing access to captive coyotes; recording work was approved by the NWRC IACUC. SW thanks Mariana Olsen for assistance with data collection, and Yellowstone National Park for permission to record. ECD thanks Amanda and Gavin McDowell from Colong Station, NSW, Australia for permission to record their animals.

References

Baker MC and Logue DM. 2003. Population differentiation in a complex bird sound: A comparison of three bioacoustical analysis procedures. Ethology 109(3):223-42.

Bergman TJ and Beehner JC. 2015. Measuring social complexity. Anim Behav 103:203-9.

Briefer E, Osiejuk TS, Rybak F, Aubin T. 2010. Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. J Theor Biol 262(1):151-64.

Briefer E, Aubin T, Lehongre K, Rybak F. 2008. How to identify dear enemies: The group signature in the complex song of the skylark *Alauda arvensis*. J Exp Biol 211(3):317-26.

Byers BE and Kroodsma DE. 2009. Female mate choice and songbird song repertoires. Anim Behav 77(1):13-22.

Cäsar C, Byrne RW, Hoppitt W, Young RJ, Zuberbühler K. 2012. Evidence for semantic communication in titi monkey alarm calls. Anim Behav 84(2):405-11.

Cover TM and Thomas JA. 1991. Elements of information theory. New York, NY: John Wiley & Sons, Inc.

Da Silva ML, Piqueira JRC, Vielliard JME. 2000. Using Shannon entropy on measuring the individual variability in the rufous-bellied thrush *Turdus rufiventris* vocal communication. J Theor Biol 207(1):57-64.

Darolová A, Krištofík J, Hoi H, Wink M. 2012. Song complexity in male marsh warblers: Does it reflect male quality? Journal of Ornithology 153(2):431-9.

Demartsev V, Kershenbaum A, Ilany A, Barocas A, Ziv EB, Koren L, Geffen E. 2014. Male hyraxes increase song complexity and duration in the presence of alert individuals. Behav Ecol 25:1451-8.

Downie JS. 2003. Music information retrieval. Annual Review of Information Science and Technology 37(1):295-340.

Doyle LR, McCowan B, Hanser SF, Chyba C, Bucci T, Blue JE. 2008. Applicability of information theory to the quantification of responses to anthropogenic noise by southeast Alaskan humpback whales. Entropy 10(2):33-46.

Edmonds B. 1999. What is complexity? The philosophy of complexity per se with application to some examples in evolution. In: The evolution of complexity. Heylighen F and Aerts D, editors. Dordrecht: Kluwer.

Ferrer-i-Cancho R. 2005. The variation of Zipf's law in human language. The European Physical Journal B-Condensed Matter and Complex Systems 44(2):249-57.

Ferrer-i-Cancho R and McCowan B. 2009. A law of word meaning in dolphin whistle types. Entropy 11(4):688-701.

Ferrer-i-Cancho R and Solé RV. 2003. Least effort and the origins of scaling in human language. Proc Natl Acad Sci USA 100(3):788-91.

Freeberg TM, Dunbar RIM, Ord TJ. 2012. Social complexity as a proximate and ultimate factor in communicative complexity. Philos Trans R Soc Lond B Biol Sci 367(1597):1785-801.

Freeberg TM. 2008. Complexity in the chick-a-dee call of carolina chickadees (*Poecile carolinensis*): Associations of context and signaler behavior to call structure. Auk 125(4):896-907.

Freeberg TM and Krams I. 2015. Does social complexity link vocal complexity and cooperation? Journal of Ornithology 156:1-8.

Fripp D, Owen C, Quintana-Rizzo E, Shapiro A, Buckstaff K, Jankowski K, Wells R, Tyack P. 2005. Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members. Anim Cogn 8(1):17-26.

Gammon DE. 2014. Seasonal patterns of vocal mimicry in northern mockingbirds *Mimus polyglottos*. J Avian Biol 45(6):545-50.

Garland EC, Noad MJ, Goldizen AW, Lilley MS, Rekdahl ML, Garrigue C, Constantine R, Hauser ND, Poole MM, Robbins J. 2013. Quantifying humpback whale song sequences to understand the dynamics of song exchange at the ocean basin scale. J Acoust Soc Am 133:560-9.

Gazda SK, Connor RC, Edgar RK, Cox F. 2005. A division of labour with role specialization in group-hunting bottlenose dolphins (*Tursiops truncatus*) off Cedar Key, Florida. Proc Biol Sci 272(1559):135-40.

Gustison ML and Bergman TJ. 2016. Vocal complexity influences female responses to gelada male calls. Scientific Reports 6:19680.

Gwilliam J, Charrier I, Harcourt RG. 2008. Vocal identity and species recognition in male Australian sea lions, *Neophoca cinerea*. J Exp Biol 211(Pt 14):2288-95.

Hallberg KI. 2007. Information in a long-distance vocal signal: Chorus howling in the coyote (*Canis latrans*). PhD thesis. The Ohio State University.

Hiebert SM, Stoddard PK, Arcese P. 1989. Repertoire size, territory acquisition and reproductive success in the song sparrow. Anim Behav 37:266-73.

Janik VM, Todt D, Dehnhardt G. 1994. Signature whistle variations in a bottlenosed dolphin, *Tursiops truncatus*. Behav Ecol Sociobiol 35(4):243-8.

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

Kershenbaum A, Sayigh LS, Janik VM. 2013. The encoding of individual identity in dolphin signature whistles: How much information is needed? PLoS One 8(10):e77671.

Kershenbaum A, Ilany A, Blaustein L, Geffen E. 2012. Syntactic structure and geographical dialects in the songs of male rock hyraxes. Proc R Soc Lond B Biol Sci 279(1740):2974-81.

Kershenbaum A, Bowles AE, Freeberg TM, Jin DZ, Lameira AR, Bohn K. 2014. Animal vocal sequences: Not the Markov chains we thought they were. Proceedings of the Royal Society B: Biological Sciences 281(1792):20141370.

Kershenbaum A, Root-Gutteridge H, Habib B, Koler-Matznick J, Mitchell B, Palacios V, Waller S. 2016. Disentangling canid howls across multiple species and subspecies: Structure in a complex communication channel. Behav Processes 124:149-57.

King SL and Janik VM. 2015. Come dine with me: Food-associated social signalling in wild bottlenose dolphins (*Tursiops truncatus*). Animal Cognition 18:1-6.

Kipper S, Mundry R, Sommer C, Hultsch H, Todt D. 2006. Song repertoire size is correlated with body measures and arrival date in common nightingales, *Luscinia megarhynchos*. Anim Behav 71(1):211-7.

Krams I, Krama T, Freeberg TM, Kullberg C, Lucas JR. 2012. Linking social complexity and vocal complexity: A parid perspective. Philosophical Transactions of the Royal Society B: Biological Sciences 367(1597):1879-91.

Larson KA. 2004. Advertisement call complexity in northern leopard frogs, *Rana pipiens*. Copeia 2004(3):676-82.

Leitão A, Ten Cate C, Riebel K. 2006. Within-song complexity in a songbird is meaningful to both male and female receivers. Anim Behav 71(6):1289-96.

Marler P and Slabbekoorn HW. 2004. Nature's music: The science of birdsong. Academic Press.

May-Collado LJ and Wartzok D. 2008. A comparison of bottlenose dolphin whistles in the Atlantic Ocean: Factors promoting whistle variation. J Mammal 89(5):1229-40.

McCowan B, Doyle LR, Hanser SF. 2002. Using information theory to assess the diversity, complexity, and development of communicative repertoires. J Comp Psychol 116(2):166-72.

Mech LD and Boitani L. 2010. Wolves: Behavior, ecology, and conservation. University of Chicago Press.

Mech LD, Smith DW, MacNulty DR. 2015. Wolves on the hunt: The behavior of wolves hunting wild prey. University of Chicago Press.

Montemurro MA and Zanette DH. 2015. Complexity and universality in the long-range order of words. arXiv Preprint arXiv:1503.01129 .

Müllensiefen D and Frieler K. 2004. Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. Computing in Musicology 13:147-76.

Nakagawa S, Waas JR, Miyazaki M. 2001. Heart rate changes reveal that little blue penguin chicks (*Eudyptula minor*) can use vocal signatures to discriminate familiar from unfamiliar chicks. Behav Ecol Sociobiol 50(2):180-8.

Narins PM and Capranica RR. 1978. Communicative significance of the two-note call of the treefrog *Eleutherodactylus coqui*. Journal of Comparative Physiology 127(1):1-9.

Nash JS and Bowles AE. 2011. N3 call types produced long-term by a killer whale of the northern resident community under controlled conditions: Characteristics, variation, and behavioral context. J Acoust Soc Am 130(4):2358.

Nowicki S, Searcy W, Peters S. 2002. Brain development, song learning and mate choice in birds: A review and experimental test of the" nutritional stress hypothesis". Journal of Comparative Physiology A 188(11-12):1003-14.

O'Connell-Rodwell CE, Wood JD, Kinzley C, Rodwell TC, Poole JH, Puria S. 2007. Wild African elephants (*Loxodonta africana*) discriminate between familiar and unfamiliar conspecific seismic alarm calls. J Acoust Soc Am 122:823.

Ord TJ, Garcia-Porta J, Ord TJ, Garcia-Porta J. 2012. Is sociality required for the evolution of communicative complexity? Evidence weighed against alternative hypotheses in diverse taxonomic groups. Philos Trans R Soc Lond B Biol Sci 367(1597):1811-28.

Palacios V, Font E, Márquez R, Carazo P. 2015. Recognition of familiarity on the basis of howls: A playback experiment in a captive group of wolves. Behaviour 152(5):593-614.

Palacios V, Font E, Márquez R. 2007. Iberian wolf howls: Acoustic structure, individual variation, and a comparison with north American populations. J Mammal 88(3):606-13.

Parsons D and Levin B. 1975. The directory of tunes and musical themes. Cambridge: S. Brown.

Pauws S. 2002. Cuby hum: A fully operational query-by-humming system. ISMIR 2002 conference proceedings, IRCAM, 2002.

Pfaff JA, Zanette L, MacDougall-Shackleton SA, MacDougall-Shackleton EA. 2007. Song repertoire size varies with HVC volume and is indicative of male quality in song sparrows (*Melospiza melodia*). Proceedings of the Royal Society B: Biological Sciences 274(1621):2035-40.

Pitcher BJ, Harcourt RG, Charrier I. 2012. Individual identity encoding and environmental constraints in vocal recognition of pups by Australian sea lion mothers. Anim Behav 83(3):681-90.

Pollard KA and Blumstein DT. 2012. Evolving communicative complexity: Insights from rodents and beyond. Philos Trans R Soc Lond B Biol Sci 367(1597):1869-78.

Quick NJ and Janik VM. 2012. Bottlenose dolphins exchange signature whistles when meeting at sea. Proc R Soc Lond B Biol Sci 279(1738):2539-45.

Radford A. 2005. Group-specific vocal signatures and neighbour–stranger discrimination in the cooperatively breeding green woodhoopoe. Anim Behav 70(5):1227-34.

Root-Gutteridge H, Bencsik M, Chebli M, Gentle LK, Terrell-Nield C, Bourit A, Yarnell RW. 2014. Identifying individual wild eastern grey wolves (*Canis lupus lycaon*) using fundamental frequency and amplitude of howls. Bioacoustics 23:55-66.

Sayigh LS, Tyack PL, Wells RS, Solow AR, Scott MD, Irvine AB. 1999. Individual recognition in wild bottlenose dolphins: A field test using playback experiments. Anim Behav 57(1):41-50.

Schlenker P, Chemla E, Cäsar C, Ryder R, Zuberbühler K. in press. Titi semantics: Context and meaning in titi monkey call sequences. Natural Language and Linguistic Theory 35(1):271-298.

Searcy WA. 1992. Song repertoire and mate choice in birds. Am Zool 32(1):71-80.

Seyfarth RM and Cheney DL. 2014. The evolution of language from social cognition. Curr Opin Neurobiol 28:5-9.

Stoica P and Moses RL. 2005. Spectral analysis of signals. Upper Saddle River, NJ: Pearson/Prentice Hall.

Suzuki R, Buck JR, Tyack PL. 2005. The use of Zipf's law in animal communication analysis. Anim Behav 69(1):9-17.

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. 2000. A procedure for an automated measurement of song similarity. Anim Behav 59(6):1167-76.

Tooze Z, Harrington F, Fentress J. 1990. Individually distinct vocalizations in timber wolves, *Canis lupus*. Anim Behav 40(4):723-30.

Wells RS. 2003. Dolphin social complexity: Lessons from long-term study and life history. In: Animal social complexity: Intelligence, culture, and individualized societies. de Waal FBM and Tyack PL, editors. Harvard University Press.

Zaccaroni M, Passilongo D, Buccianti A, Dessi-Fulgheri F, Facchini C, Gazzola A, Maggini I, Apollonio M. 2012. Group specific vocal signature in free-ranging wolf packs. Ethol Ecol Evol 24(4):322-31.

Zuberbühler K. 2002. A syntactic rule in forest monkey communication. Anim Behav 63(2):293-9.

FIGURES

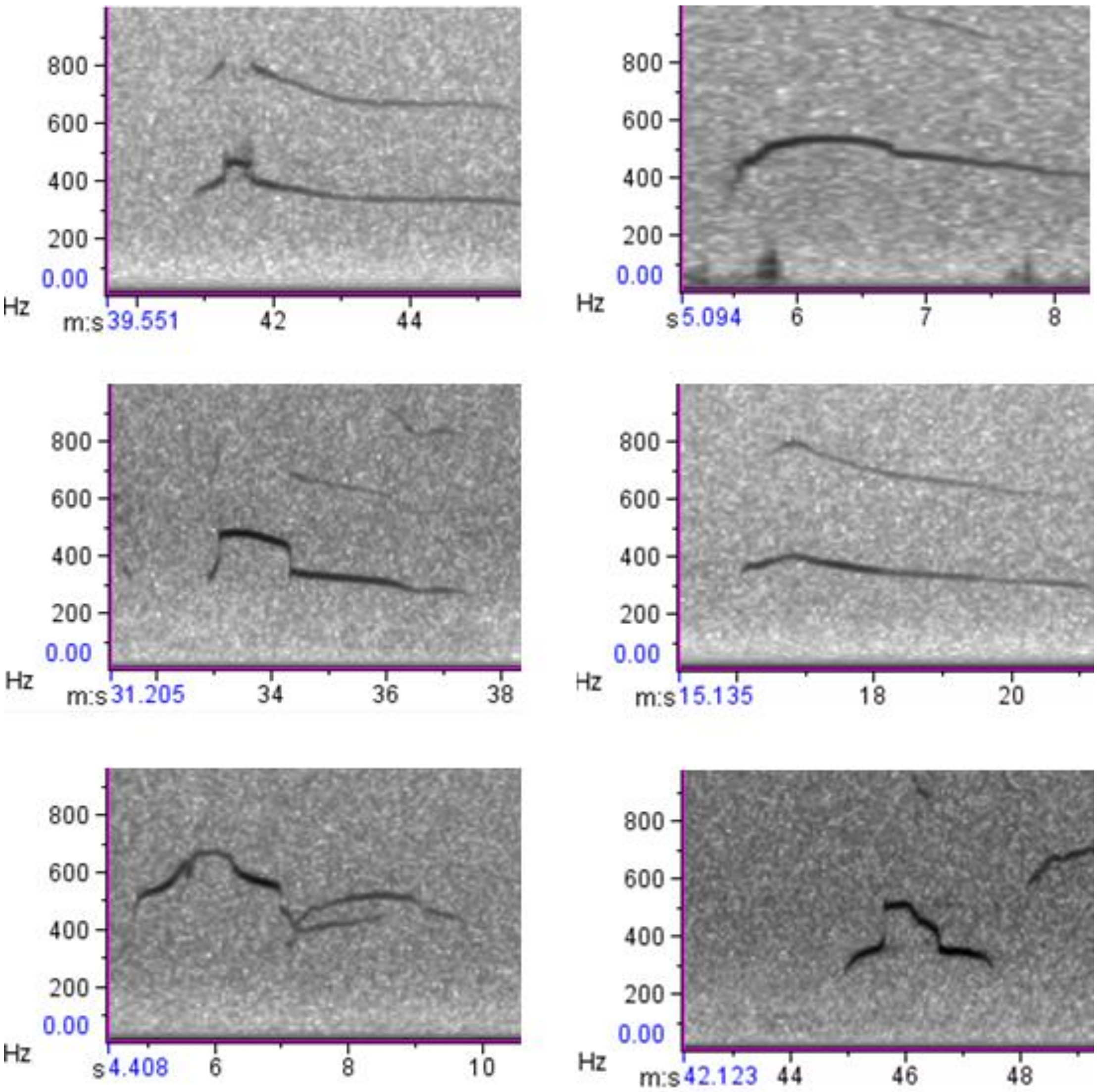

Figure 1. Six examples of wolf howls, with time on the x-axis and frequency on the y-axis. All howls show some pattern of rising and falling frequencies, but the determination of which howl is most complex appears wholly subjective.

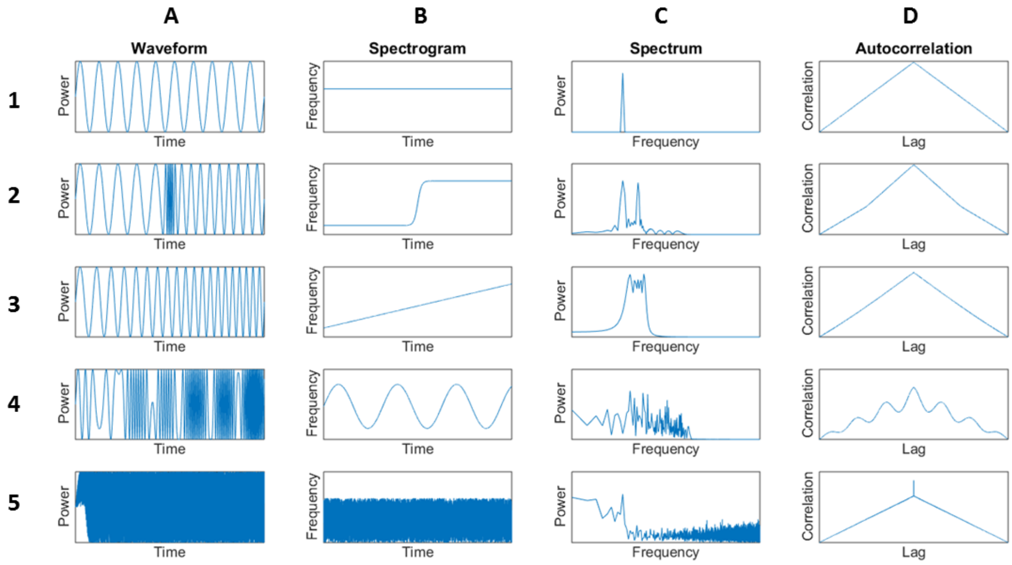

Figure 2. Simulated sounds demonstrating the relationship between waveform, spectrogram, spectrum, and autocorrelation. The first column (A) shows five different waveforms: (1) constant frequency sine wave; (2) rapid doubling in frequency; (3) constantly increasing in frequency; (4) oscillating frequency; and (5) random waveform. The second column (B) shows the frequency of the waveform with time. Column (C) shows overall spectra for these waveforms: a single peak where one frequency is present (1), two peaks where two frequencies exist (2), a range of frequencies in (3) and (4), and all frequencies present in the case of white noise (5). Column (D) shows the autocorrelation of the spectrogram.

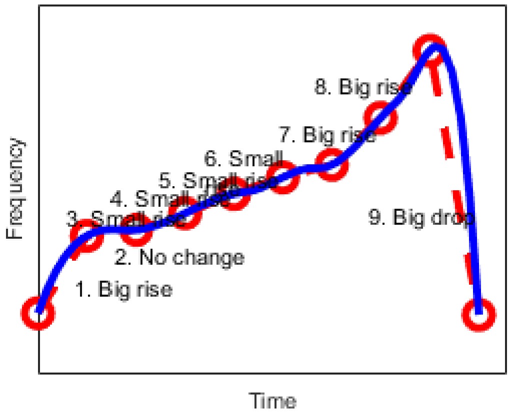

Figure 3: Example of a Parsons code representation of a vocal signal (blue). The time-course is divided into ten equal sections (red), and the frequency change for each section is recorded only as "big rise", "small rise", "no change", "small drop", or "big drop".

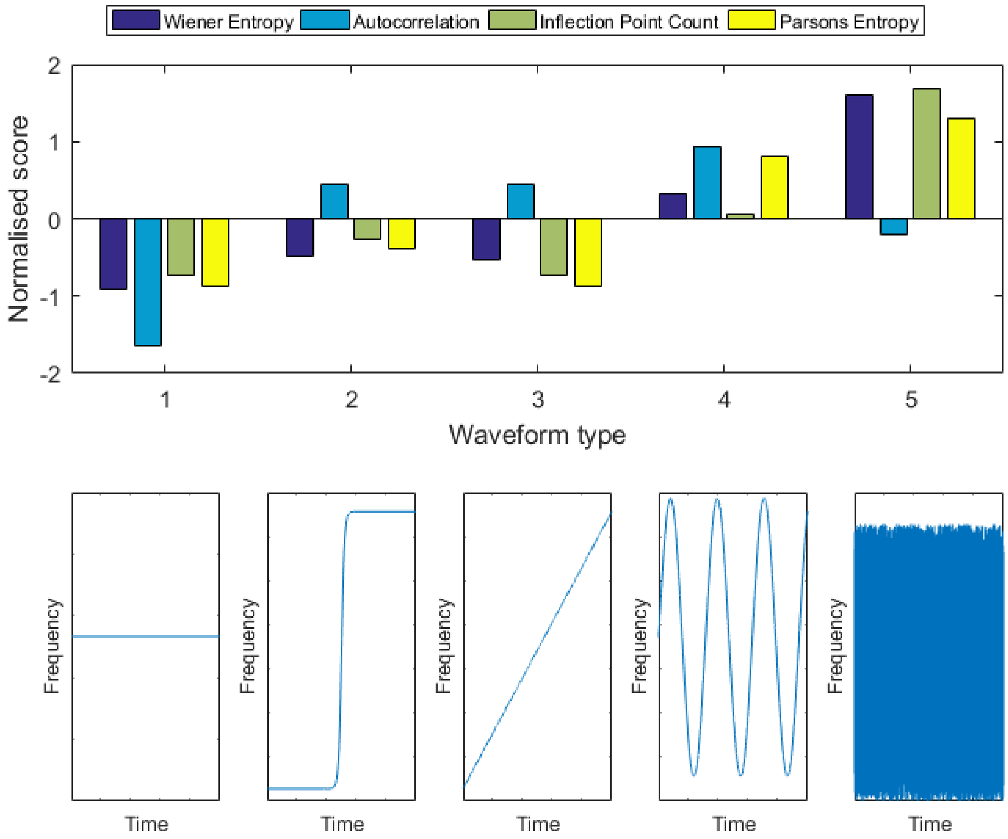

Figure 4. Complexity scores for the four metrics, for each of the five waveforms shown in Figure 2. The bar charts are normalised for comparison by subtracting the mean and dividing by the standard deviation.

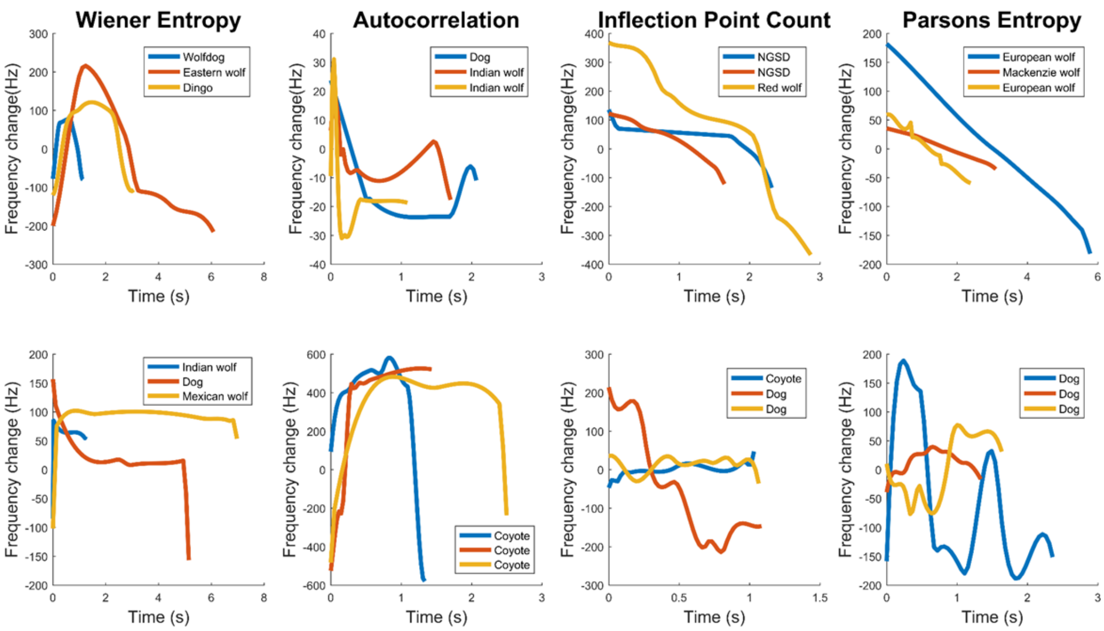

Figure 5. Examples of howls with low metric values (top row), and high metric values (bottom row), for each of the four metric types. Each plot shows those three sample howls with the highest or lowest values for each particular metric (not necessarily the same sample howls for each metric). The y-axis indicates frequency deviation from the median value of each howl, to allow a clear comparison between howls of differing frequencies.

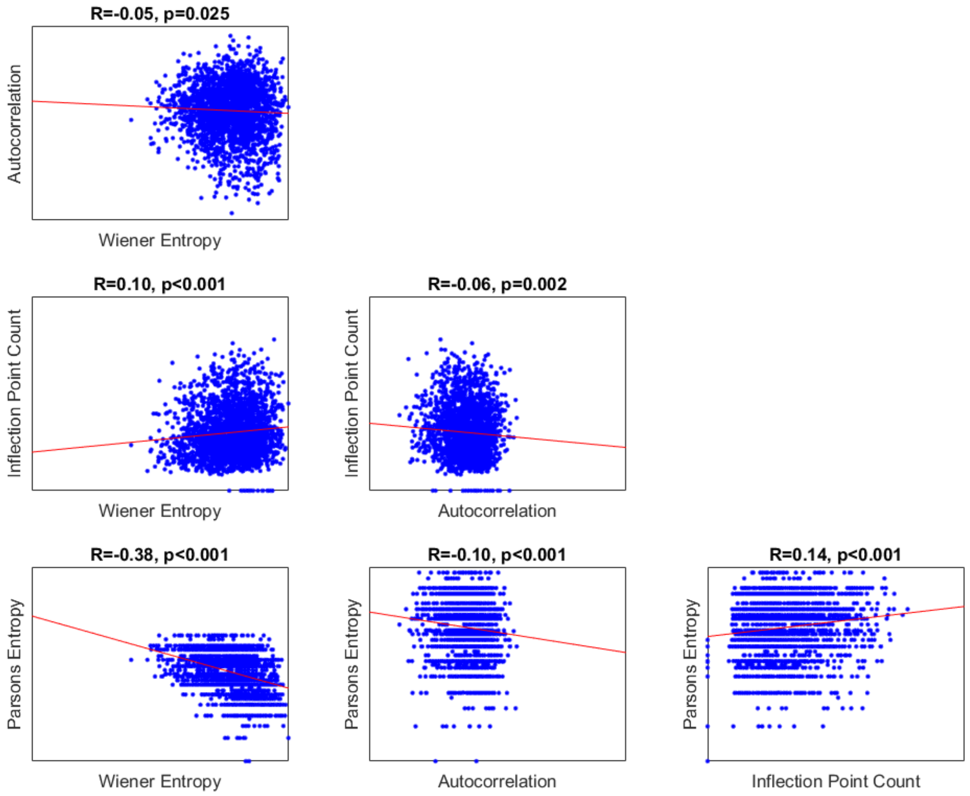

Figure 6. Pairwise correlation between each of the metrics. The red line indicates line of best fit, and *R* and *p* values given in the title of each plot are for Pearson correlation.