Animal Behaviour 112 (2016) 39–51

Contents lists available at ScienceDirect

Animal Behaviour

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

Commentary

Automated analysis of song structure in complex birdsongs

Mareile Große Ruse [a,*], Dennis Hasselquist [b], Bengt Hansson [b], Maja Tarka [c], Maria Sandsten [d]

[a] Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark
[b] Department of Biology, Lund University, Lund, Sweden
[c] Center for Biodiversity Dynamics (CBD), Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[d] Centre for Mathematical Sciences, Lund University, Lund, Sweden

ARTICLE INFO

Article history:
Received 23 June 2015
Initial acceptance 20 July 2015
Final acceptance 13 October 2015
Available online
MS. number: 15-00536R

Keywords:
ambiguity spectrum
automated song recognition
birdsong
clustering
great reed warbler
multitaper
song analysis
syllable detection

Understanding communication and signalling has long been strived for in studies of animal behaviour. Many songbirds have a variable and complex song, closely connected to territory defence and reproductive success. However, the quantification of such variable song is challenging. In this paper, we present a novel, automated method for detection and classification of syllables in birdsong. The method provides a tool for pairwise comparison of syllables with the aim of grouping them in terms of their similarity. This allows analyses such as (1) determining repertoire size within an individual, (2) comparing song similarity between individuals within as well as between populations of a species and (3) comparing songs of different species (e.g. for species recognition). Our method is based on a particular feature representation of song units (syllables) which ensures invariance to shifts in time, frequency and amplitude. Using a single song from a great reed warbler, *Acrocephalus arundinaceus*, recorded in the wild, the proposed algorithm is evaluated by means of comparison to manual auditory and visual (spectrogram) song investigation by a human expert and to standard song analysis methods. Our birdsong analysis approach conforms well to manual classification and, moreover, outperforms the hitherto widely used methods based on mel-frequency cepstral coefficients and spectrogram cross-correlation. Thus, our algorithm is a methodological step forward for analyses of song (syllable) repertoires of birds singing with high complexity.

Birdsong is among the most prominent and widespread avian behaviours, dominating the audial environment in spring and early summer. Considerable research effort has been devoted to questions such as how birds sing and why birds have such elaborate songs (Catchpole & Slater, 1995; Miller & Kroodsma, 1996). Central to most research questions involving birdsong is the need for classifying, comparing and quantifying sounds in the song within and among individuals. Typical questions include how song repertoire size or song variability influence mate choice and male–male competition (e.g. Hasselquist, Bensch, & von Schantz, 1996; Horn & Falls, 1996; Searcy & Yasukawa, 1996). This involves comparisons of songs within an individual to estimate song complexity (Catchpole, 1976), comparing song similarity among individuals within a species (e.g. neighbour song matching, Falls, 1985; Horn & Falls, 1988) and variation between individuals within a population (Slater, Clements, & Goodfellow, 1984) and over time (Lehtonen, 1983), investigating geographical variation (Catchpole & Rowell, 1993) and song dialects (Espmark, Lampe, & Bjerke, 1989; McGregor, 1980; Mundinger, 1980), as well as variation between species, allowing species recognition (Kreutzer & Güttinger, 1991; Martens, 1996; Miller, 1996). Studies assessing vocal development and song learning endeavour to quantify similarities (imitation) between the song of a young bird and its tutor (Kroodsma & Konishi, 1991; Nottebohm, 1991; Slater & Ince, 1982). Classification of song sounds and estimation of song complexity (i.e. song repertoire size) can be conducted either (1) on whole song strophes in species with low to medium song complexity such as the chaffinch, *Fringilla coelebs* (Slater, 1983) and the American redstart, *Setophaga ruticilla* (Lemon, Cotter, MacNally, & Monette, 1985) or (2) on smaller sound entities, such as syllables, which are discrete sound units that build up (often a large number of different) song strophes in species with higher song complexity such as the great reed warbler (GRW), *Acrocephalus arundinaceus* (Hasselquist, 1998; Hasselquist et al., 1996; Węgrzyn & Leniowski, 2010) and the pied

* Correspondence: M. Große Ruse, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark.
E-mail address: mareile@math.ku.dk (M. Große Ruse).

http://dx.doi.org/10.1016/j.anbehav.2015.11.013
0003-3472/

flycatcher, *Ficedula hypoleuca* (Eriksen, Slagsvold, & Lampe, 2011; Lampe & Espmark, 2003).

The hitherto standard methods to classify song entities (syllables) has been by means of the audial and visual comparison of syllables (see e.g. Catchpole, 1976; Hasselquist et al., 1996; Węgrzyn, Leniowski, & Osiejuk, 2010), where the latter is often conducted based on syllable spectrograms (Adret, Meliza, & Margoliash, 2012; Węgrzyn & Leniowski, 2010). Unfortunately, these approaches are often time consuming, prone to observer bias and subjectivity, non-numerical (making statistical analyses problematic) and perform less well on songs with large syllable repertoires or with complex structures of song strophes/syllables (Clark, Marler, & Beeman, 1987; Williams, 1993; Williams & Slater, 1991). The algorithm we propose enables an automated objective classification of birdsongs and thereby facilitates the assessment of the song repertoire of an individual bird and its temporal development, as well as comparisons of song structures among birds within and between different populations.

Automated song analysis is typically conducted by subdividing a song into smaller entities (e.g. syllables or even larger song sections as in Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000) and representing them in terms of features, i.e. a collection of characterizing properties. These features can be intuitive parameters such as the time length of a unit, its power or pitch or more involved quantities, such as mel-frequency cepstral coefficients (MFCC) or wavelet coefficients (see below). The selection of features, however, should be guided by the purpose of the study, properties of the data and not least by available computational resources. The feature representation summarizes important characteristics of the song units, which facilitates a comparative analysis in terms of a similarity measure that operates on the feature basis. If the differences of interest between syllables are known in advance and are sufficiently pronounced, the representing features should be chosen such that they are able to reflect these differences prominently. If, for example, one factor of interest is a syllable's length, the time duration will be the natural choice and should be included in the feature representation. However, if there is no prior knowledge of the underlying factors for classification or if syllable characteristics are very complex and not straightforward to capture, the purported feature space should be sufficiently rich in order to facilitate detection of various syllable characteristics. Methods for song analysis that have been proposed in the literature differ mainly in the way song units are chosen, which features are extracted and how similarity between features of song units is assessed. The choice of small entities such as single syllables as the basic building blocks facilitates song complexity analysis of songs with highly variable strophes, such as those of most European and Asian warblers, flycatchers, thrushes, wrens and chats. Moreover, sections of a recording that are affected by substantial background noise can easily be discarded, which proves beneficial if only field recordings are available. If, however, the whole recording is confounded by noise, it may be hard to tell subsequent syllables apart, which hampers an unambiguous definition/detection of syllables.

The most well-known technique in the context of birdsong analysis is the cross-correlation approach applied to the spectrogram (SPCC; Keen, Ross, Griffiths, Lanzone, & Farnsworth, 2014). This method, however, suffers from sensitivity to natural jitters of components at time and frequency locations as well as sensitivity to noise. Tchernichovski, Lints, Mitra, and Nottebohm (1999) therefore applied the more robust multitaper (MT) spectrograms for syllable-based feature extraction. Kogan and Margoliash (1998) showed that spectrogram-based features are inferior to a syllable representation by means of MFCC when recordings are substantially affected by background noise. Syllable representation by time-varying sinusoids was applied to recognition of bird species with relatively simplistic songs by Härmä (2003), while Somervuo and Härmä (2004) combined this approach with a clustering of syllables in terms of the *k*-means algorithm. In a comparative study, Somervuo, Härmä, and Fagerlund (2006) found that the sinusoidal model approach was clearly outperformed by the more complex MFCC-based syllable representation. A combination of the latter and some descriptive parameters for syllable representation was investigated by Fagerlund (2007) and by Trifa, Kirschel, Taylor, and Vallejo (2008) in the context of species recognition. For more complex birdsong syllables the comparably simple models such as the sinusoidal approach and the MFCC representation may fail to capture central information of the signals and more sophisticated representations such as wavelet decompositions have been considered (Selin, Turunen, & Tanttu, 2007). (For a recent comparison of various methods, such as SPCC, dynamic time warping and pitch-frequency analysis, see Keen et al., 2014; Meliza, Keen, & Rubenstein, 2013.) Algorithms employing some of the previously mentioned techniques have been implemented and made available as ready-to-use programs, among them Sound Analysis Pro (Tchernichovski & Mitra, 2004), Luscinia (Lachlan, 2007), Avisoft-SASLab Pro (Specht, 2004) or Praat (Boersma & Weenink, 2001).

For the within-species analysis of songs with elaborate complexity (e.g. such as those of the GRW), however, techniques with an additional level of sensitivity are required.

In this paper, we propose a fully automated method tailored to syllable-based, within-species analysis of field recordings of complex birdsongs. The algorithm, which allows an unbiased, reproducible and reliable song analysis, is a three-step procedure, comprising syllable detection, representation and comparison (with further clustering, if required). The novelties in our approach are the usage of the ambiguity spectrum (a transformation of the spectrogram and also called the Doppler-lag spectrum) for feature extraction and a novel similarity measure for subsequent syllable comparison.

The advantage of the ambiguity spectrum as opposed to the popular standard spectrogram is its invariance to time and frequency shifts of syllables, as the ambiguity spectrum is always centred at zero Doppler frequency and time lag (Boashash, 2003). It therefore focuses solely on the relations between different time and frequency components in a syllable. As already successfully employed by, for example, Meliza et al. (2013) and Tchernichovski et al. (1999), as well as in our preliminary study (Sandsten, Tarka, Caissy-Martineau, Hansson, & Hasselquist, 2011), we used MTs for noise-robust spectrogram estimation. Syllable clustering is achieved by means of a hierarchical clustering algorithm where the number of clusters is objectively estimated by the Silhouette quality criterion. The proposed method, as well as comparative established approaches based on MFCC and SPCC, is evaluated on a real data set by comparison to a 'ground truth' given by a human expert (D.H.) with long experience in GRW song syllable analysis (Hasselquist, 1998; Hasselquist et al., 1996). All computations were conducted in MATLAB (MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, MA, U.S.A.) and we can provide a ready-to-use code upon request.

METHODS

Evaluation of the algorithm by comparison to human expert clustering and to methods based on MFCC and SPCC is conducted by means of three examples of increasing complexity. The underlying data represent the type of data typically obtained in the challenging setting of field recordings of complex birdsongs. In a first example, we use our algorithm and two alternative approaches (MFCC and SPCC) to recover a classification of syllables into two visually well-separated groups. The second example generalizes this setting to a