

Automated annotation of Orthoptera songs: first results from analysing the DORSA sound repository

KLAUS RIEDE, FRANK NISCHK, CHRISTIAN DIETRICH, CHRISTIAN THIEL AND FRIEDHELM SCHWENKER

(KR) Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany. Email: k.riede.zfmk@uni-bonn.de
 (FN) Department of Neural Information Processing, University of Ulm, Germany. Email: franknischk@aol.com
 (CD, CT, FS) Brüsseler Str. 63, D-50672 Köln, Germany. Email: {christian.dietrich, christian.thiel, friedhelm.schwenker}@uni-ulm.de

Abstract

Orthoptera songs are widely used for the description and diagnosis of new species. Most of the corresponding sound recordings are in analogue format (tapes), widely scattered among institutions, and only a small fraction is accessible as an organized collection ('phonothek'). Approximately 12,000 Orthoptera sound recordings, representing about 4,000 species from all biogeographic regions, have been digitized and stored in a database during the DORSA project (Digital Orthoptera Specimen Access – www.dorsa.de). Together with images and collection data of voucher specimens, the DORSA serves as a 'Virtual Museum', summarizing distributed collections and phonotheks from several German researchers and institutions. A subset of recordings was used to develop automatic sound recognition tools, by using neural networks fed by acoustic parameters. Relevant parameters, such as carrier frequency and pulse repetition rate, were determined by a special software module, which could then be used to extract those features from all cricket songs hitherto available in the DORSA database. These parameters were then added as annotations to the individual song recordings within the database. The enriched database tables were analysed, revealing outliers due to low-quality recordings or misidentification, permitting a cleaning of the data. For recordings from a limited geographic range, pulse intervals and carrier frequency are sufficient to identify plausible matchings between archived songs and new sound recordings.

Key words

bioacoustics, Grylloidea, song parameters, species identification, artificial neural networks, databases

Abstract (French)

Annotation automatisée de chants d'Orthoptères: premiers résultats après l'analyse du dépôt d'archives du son DORSA.

Les chants d'Orthoptères ont été beaucoup utilisés pour décrire et déterminer de "nouvelles" espèces. La plupart des enregistrements de son sont sous format analogique (cassettes) dispersés dans diverses institutions et seule une petite partie est accessible sous forme de collection structurée (phonothèque). Plus de 12 000 enregistrements du son d'Orthoptères, représentant environ 4,000 espèces de toutes les régions bio-geographiques, ont été digitalisés et mises dans une base de données au cours du projet DORSA (Digital Orthoptera Specimen Access – www.dorsa.de). Ce dépôt d'archives de son est maintenant accessible on-line à l'aide de la base de données SYSTAX. En joignant des illustrations aux données de collection des spécimens de référence, DORSA peut servir de "Musée virtuel", qui regroupe des collections et des phonothèques de nombreux chercheurs et institutions allemands. Un sous-ensemble d'enregistrements a permis de développer des logiciels de reconnaissance de son automatique, en utilisant un réseau neuronal combine avec une extraction des paramètres acoustiques. Le logiciel a pu déterminer que les paramètres les plus importantes sont les taux de répétition des impulsions et la fréquence porteuse. Ces paramètres

ont alors été ajoutés aux enregistrements individuels dans la base des données. Néanmoins, cette base de données annotée peut dès à présent servir de référence qui peut être rapidement consulté pour les paramètres de chants importants, tout en identifiant des sous-ensembles de chants qui probablement se révéleront être semblables une fois que de nouveaux enregistrements de son seront disponibles.

Introduction

The species-specific, stereotyped songs of male Orthoptera can be used as a highly reliable feature for the recognition and description of species (Otte 1994). This requires methods for objective song description, some of which were developed long before present-day advanced technologies for sound recording and analysis. In an enlightening review, Ragge & Reynolds (1998) outline the history of song representations, including a reproduction of A. Yersin's attempt at musical notation for the songs of European Orthoptera, dating from 1854 (*loc. cit.*, Fig. 34). During the 20th century, the development of tape recorders, oscilloscopes and spectrographs resulted in the well-known graphical representations used in modern bioacoustics. In addition, several Orthoptera species became model organisms for neuroethology, revealing the underlying physiological processes of song production, hearing and phonotaxis (Huber *et al* 1989, Schildberger 1994, Helversen & Helversen 1998, Römer & Krusch 2000). Taxonomic and behavioral studies were successfully combined, and resulted in vast collections of sound recordings. Ideally, they were deposited in major sound archives, such as the Wildlife Section of the National Sound Archive (NSA) in Britain, as is the case for Ragge's extensive song recordings of Western European grasshoppers. In Germany, which has a strong tradition in insect bioacoustics and neuroethology, recordings were not centrally collected, but remained with the respective researcher or affiliated institution (*e.g.*, Faber 1953).

In an attempt to safeguard these distinct tape collections and phonotheks, songs were digitized, put into databases, and published on the Internet as a 'Virtual phonothek' as part of the DORSA project (DORSA: Digital Orthoptera Specimen Access, see <http://www.dorsa.de> and Ingrisch *et al.* 2004). A subset of this digitized sound repository was used to develop automated sound classification tools using neural networks (Dietrich *et al.* 2004). The vision was to integrate acoustic recording into Rapid Assessment Programs, as a noninvasive technique to classify and map acoustic diversity (Riede 1993). Software tools should enable field researchers to classify sound recordings from complex acoustical environments, such as tropical rainforest canopies, by automated filtering, analysis and

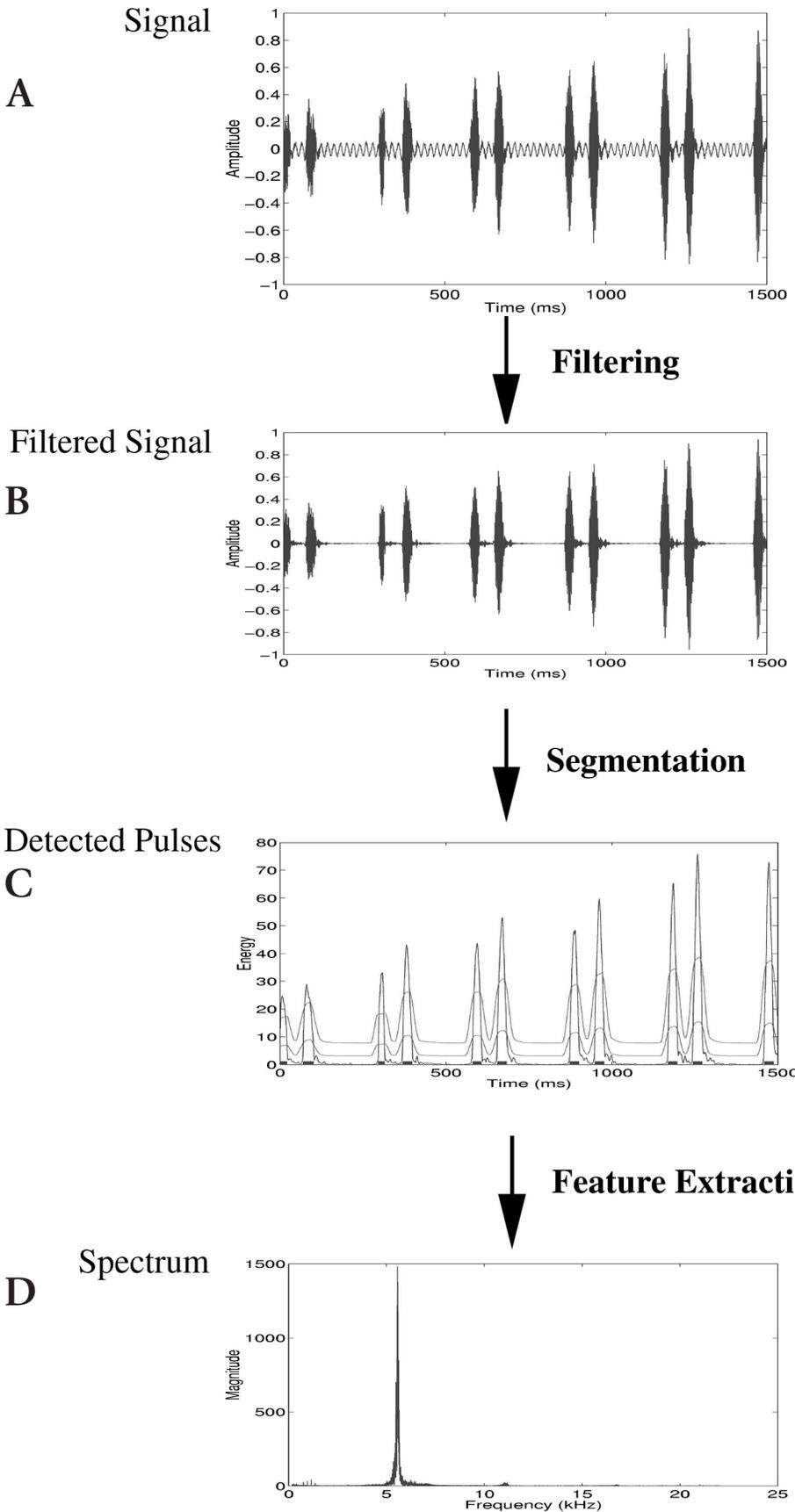


Fig. 1. Preprocessing of Orthoptera songs is illustrated using a recording of *Noctitrella glabra* Ingrisch, 1997 (Cryllidae: Podoscirtinae: Podoscirtini). A. Recording by S. Ingrisch, Surat Thani, Khao Sok, Thailand 25.1.1997). B. After amplitude normalisation, the sound is band-pass filtered to suppress environmental noise. C. Two non-constant threshold functions (lower thin lines) are calculated from the local energy of the signal. Based on these threshold functions, the position and duration of the pulses is determined [indicated through the bars at the abscissa and the energy peaks (solid lines)]. Position and duration of the pulses are highly discriminative, particularly in combination with spectral features, D.

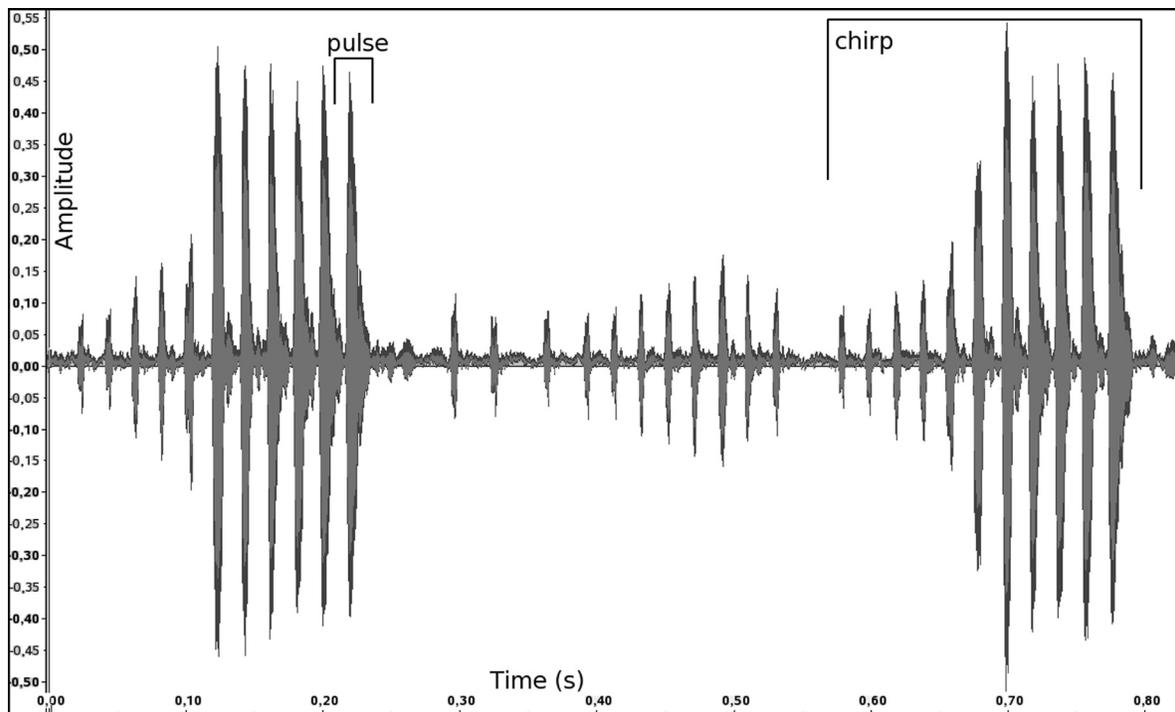


Fig. 2. Amplitude-vs-time signal from an individual of the cricket species *Homoeoxipha lycoides* (Gryllidae: Trigonidiinae: Trigonidiini). Recording by S. Ingrisch, December 1998, Khumphaya, Thailand.

possibly identification of songs.

In the study reported here we used the feature-extraction modules from our automatic sound classification software (Dietrich *et al.* 2004) to determine fundamental acoustic parameters such as carrier frequency and pulse rate of all digitized cricket songs available in the DORSA database. Extracted parameters were then annotated as sound metadata. This procedure speeds up feature extraction, hitherto done manually, by orders of magnitude. We demonstrate that such automated feature extraction is reliable and feasible for crickets (Grylloidea), thereby facilitating the generation of a look-up table which can be used to navigate and query the huge DORSA sound repository. Because the locality of the original recording is known, data subsets for certain geographical regions can be extracted for decentralized use by field researchers. Possible applications are the rapid identification and mapping of selected indicator species, or the diagnosis of unknown songs, which are strong indicators for undescribed, 'new' species within a certain area.

Methods

Orthoptera song repository within DORSA (Digital Orthoptera Specimen Access).—The DORSA multimedia database comprises 30,000 images of type specimens and 12,000 sound files. These are linked to voucher specimens with collection data, representing 16,000 specimen records from approximately 4,000 species. The sound files were provided in digital (wav) format by the authors for the DORSA project (www.dorsa.de, see Ingrisch *et al.* 2004). Most sound files are fully accessible through the SYSTAX database (<http://www.biologie.uni-ulm.de/systax/>) and can be used noncommercially, citing the source and the recordist. In addition, DORSA specimen information is reciprocally linked to the Orthoptera Species File (OSF – <http://osf2x.orthoptera.org/OSF2.3/>) accessible through the Global Biodiversity Information Facility (www.gbif.org).

Toolkit for Orthoptera Song Recognition and Analysis (TOSRA).—TOSRA consists of several modules implemented in C/C++ and MATLAB. They were developed to classify Orthoptera songs, based on neural and statistical pattern recognition algorithms, described in detail by Dietrich *et al.* (2004) and Schwenker *et al.* (2003). The first step of the insect song processing is resampling the sound files to a standard sampling rate of 44.1 kHz and normalising of the signal to its maximal amplitude, to suppress the influence of the sound volume. Environmental noise reduction is then performed by band-pass filtering (see Fig. 1). Details on the filtering procedure are given in Dietrich *et al.* (2004). Because temporal structure is among the most prominent features to differentiate insect songs at species level, pulse detection is a crucial part of the automated song classification, which is accomplished in the next step.

As shown in Fig. 1, the TOSRA pulse-detection algorithm is based on a lower and an upper threshold function, derived from the local sound energy. To take into account variation of the energy within a single chirp (Figs 1, 2) the threshold functions must not be constant. A pulse onset is detected if the signal exceeds both the lower and upper threshold function within a short, predefined time frame. The pulse offset is determined analogously. These two functions allow determination of the exact temporal position of single pulses, so that the following features can be calculated for each song:

- Distance between consecutive pulses
- Pulse length
- Frequency contour of pulses
- Energy contour of pulses
- Time Encoded Signal of pulses
- Maximal amplitude of pulses.

The extracted features then serve as inputs to a Radial Basis Function (RBF) neural network classifier (Schwenker *et al.* 2001). Before the RBF net can be used for classification, it has to be trained by a

Table 1. Error rates for the automated classification of cricket and katydid songs determined for different single features: pulse distance, pulse length, pulse frequency, energy contour, Time-Encoded-signals (TES) and maximal amplitude of pulses. The classifier performance based on single features can be improved significantly by combining the single-classifier decisions into an overall decision. In particular the decision template approach is highly accurate.

| | Crickets [% error] | Katydids [% error] |
|---------------------------------|-----------------------|-----------------------|
| Pulse distance | 36 | 37 |
| Pulse length | 60 | 34 |
| Frequency contour | 73 | 80 |
| Energy contour | 60 | 61 |
| Time-Encoded-Signals | 77 | 66 |
| Max. amplitude | 81 | |
| CFT-Fusion (averaging) | 34 | 31 |
| CFT-Fusion (decision templates) | 6 | 13 |

supervised learning procedure utilizing the extracted feature vectors together with the corresponding signal-class labels (the species name of the singing insect). For this training procedure, a subset of the sound data has to be used for each species. As a test-set for automatic identification, 53 species of crickets and katydids were

selected from the DORSA database. Classification results based on a statistical cross-validation testing procedure indicate very high classification accuracy (see Table 1).

Annotation and visualisation of song parameters.—The neural network classification shown in Table 1 is based on a comparatively small subset from only 53 species, and is not applicable to the entire DORSA sound archive. However, the TOSRA preprocessing routines could be used to extract important parameters such as carrier frequency or pulse distance from all songs. Therefore, the feature extraction module was applied to all DORSA sound files classified as ‘Grylloidea’ by a batch routine, calculating ‘carrier frequency’, ‘pulse distance’, ‘pulse frequency’ or ‘duty cycle’ (see Fig. 2 for illustration of features). Results were stored within additional parameter columns of the respective soundfile database tables.

Once stored within the database, parameters can be tabulated together with specimen data (cf. Table 2), or visualised by graphics programs. Particularly useful was the application of a desktop Geographical Information System (GIS: ArcView 3.2 by ESRI) for “mapping” the parameter space. Such GIS software is designed to connect spatial data with attribute tables. In our case, song parameters are plotted within the parameter space, and each point is linked to tabulated specimen data. This allows user-friendly exploration of the parameter space, by producing distinct maps: individual species can

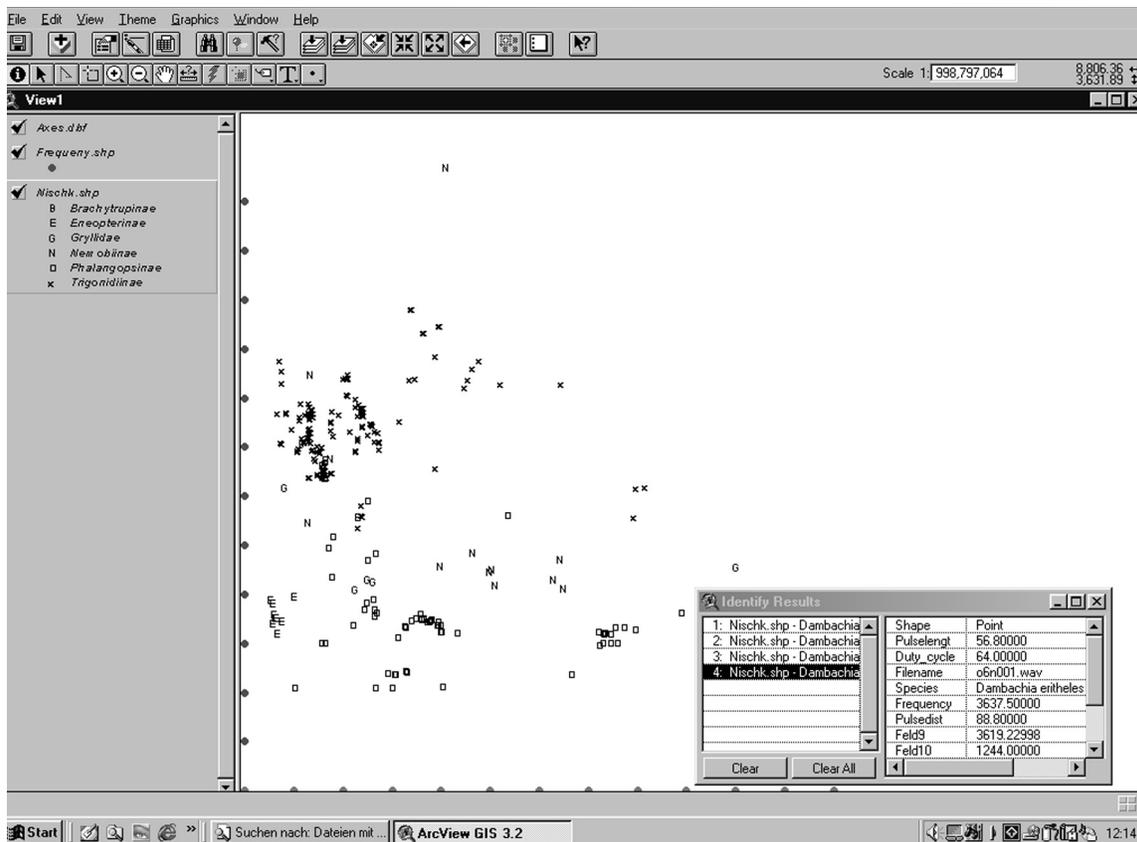


Fig. 3. Examples of parameter space with attributes. GIS software (ArcView3.2 by ESRI) used for plotting, which allows retrieval of underlying database information (species, specimen, parameters) by clicking on a certain point within the parameter space. This information then appears within a separate box. Dots scaling the axes indicate frequency intervals of 1 kHz (y-axis) and pulse distance at 10-ms intervals (x-axis). Parameter space for crickets recorded in Amazonian lowland forest, Ecuador, by Frank Nischk. Higher taxonomic units are labelled, revealing the low carrier frequencies of phalangopsine calling songs (squares). For this geographically limited subset of syntopically occurring crickets, parameters show clearly recognizable clusters coinciding with morphospecies. Therefore, pulse rate and carrier frequency can be used as a look-up table when searching for similar recordings at this site.

be highlighted, legends on the level of individual species or higher taxa can be generated 'on the fly', and outliers can be identified and inspected by clicking on the respective data points (Figs 3-5).

Results

To investigate the reliability of the automated feature extraction, some songs were analysed manually, using standard sound-analysis software. Subsequently, the resulting table was analysed for inconsistencies and outliers, and the identified 'corrupt' files were removed. Finally, the acoustic parameter space was plotted and analysed, using a Geographic Information System.

Comparison of automated vs conventional feature extraction.—For a subset of Ecuadorian trigoniid (Trigonidiinae) species, parameters were compared directly with extensive conventional sound analysis of trigoniid songs made by Nischk (1999). Table 2 shows a remarkable coincidence of results. However, Nischk (1999) describes additional, idiosyncratic 'secondary song features' such as within-pulse frequency modulations, regularity of pulse distances or grouping of pulses. Such groups are characterized as chirps, and their species-specific number of pulses can be an additional, highly reliable feature of the species song. These secondary features are 'lost' during the TOSRA simple feature extraction and representation within the parameter space (cf. Fig. 5), but they are evident at first sight in oscillograms of original wav-files, which can always be retrieved from the SYSTAX database.

Up to now, only a small subset of the voucher specimens collected

by Nischk has been described taxonomically (Desutter-Grandcolas 2000, Nischk & Otte 2000). The majority of voucher specimens was determined to subfamily or genus level. Specimens which evidently belong to a hitherto undescribed species are labelled with a preliminary code, and the respective voucher specimens can be considered as 'types of tomorrow', waiting for taxonomic description or revision. These specimens are stored in the database with their preliminary code name, (e.g., Phalangopsinae sp. PhalOtLP2, Nischk & Otte 2000, p. 248, or the Trigonidiinae listed in Table 2). Therefore, the preliminary code names are indications, but by no means equivalents of well-defined species, until a thorough taxonomic study is completed. Associated sound recordings provide additional and important bioacoustic characters, which should be used in species descriptions. For example, individuals of TrigoSP3 are morphologically undistinguishable, but fall into two groups (n and h) according to distinct pulse distances. Therefore, a closer look into the morphology might reveal subtle differences, which justify description of two species, to be easily distinguished by their distinct songs.

Data cleaning.—Species descriptions including song parameters were published by Nischk and Otte (2000) for four new genera and 10 new species of Ecuadorian Phalangopsinae. In this case, outliers were easily detected by mapping song parameters for these species with the GIS software (Fig. 3, 4). Selecting song parameters for *Hattersleya clandestina* (Nischk & Otte 2000) revealed two clusters with a similar carrier frequency, but distinct pulse distances around 16 ms and 70 ms, respectively (Fig. 4). Closer inspection of the

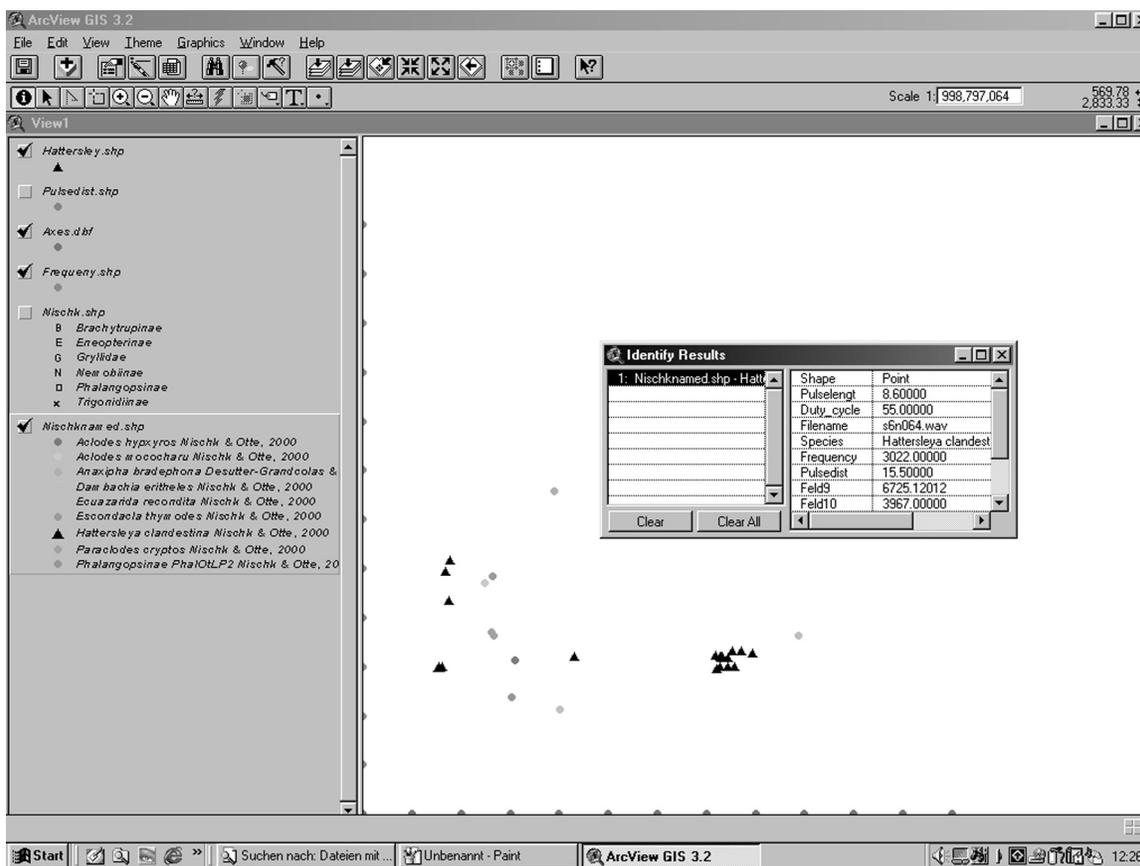


Fig 4. Examples of parameter space with attributes (contin.). Identification of outliers for *Hattersleya clandestina*: recordings for this species (labeled with black triangles) reveal two clusters. The widely scattered cluster at the left is due to low quality recordings, which consequently were removed during data cleaning (see text for further details).

Table 2. Comparison of automated feature extraction using the TOSRA module with results from conventional song analysis. Parameters obtained using automated feature extraction are labelled: A; conventional measurements of pulse distance and carrier frequency taken from the literature, are labelled: N (Nischk 1999, pp. 113 ff.). Original sound files can be retrieved through the SYSTAX Portal (<http://www.biologie.uni-ulm.de/systax/portal/index.html>) by searching for the species name or sound-file code.

| Species | Filename (wav format) | Duty cycle (A) | Pulse distance (A) [ms] | Pulse distance (N) [ms] | Carrier frequency (N) [Hz] | Carrier frequency (A) [Hz] |
|--|-----------------------------|----------------------|----------------------------------|----------------------------------|-------------------------------------|-------------------------------------|
| <i>Anaxipha bradephona</i> Desutter-Grandcolas & Nischk, 2000 | s5n012 | 64 | 39.1 | 38.1 | 6583 | 6570 |
| <i>Anaxipha tacheiphona</i> Desutter-Grandcolas & Nischk, 2000 | s6n023 | 84 | 12.5 | 10.7 | 7646 | 7674 |
| <i>Anaxipha tacheiphona</i> Desutter-Grandcolas & Nischk, 2000 | s6n029 | 88 | 13.6 | 11.2 | 7105 | 7122 |
| <i>Anaxipha tacheiphona</i> Desutter-Grandcolas & Nischk, 2000 | s7n057 | 85 | 13.4 | 11.1 | 7278 | 7303 |
| Trigonidiinae TrigoSP03h | s6n048 | 67 | 9.7 | 9.6 | 7369 | 7376 |
| Trigonidiinae TrigoSP03h | s6n062 | 71 | 8.6 | 8.6 | 7722 | 7722 |
| Trigonidiinae TrigoSP03h | s7n045 | 63 | 10.9 | 10.8 | 6915 | 6952 |
| Trigonidiinae TrigoSP03h | s6n082 | 66 | 11.2 | 11 | 6964 | 6981 |
| Trigonidiinae TrigoSP03n | s7n024 | 64 | 14 | 13.8 | 6978 | 6932 |
| Trigonidiinae TrigoSP03n | s7n007 | 66 | 13.4 | 13.4 | 7253 | 7223 |
| Trigonidiinae TrigoSP03n | s7n081 | 63 | 15.1 | 14.9 | 6467 | 6434 |
| Trigonidiinae TrigoSP04 | s7n092 | 40 | 52.2 | 52 | 8311 | 8294 |
| Trigonidiinae TrigoSP04 | s7n040 | 49 | 39.1 | 39.1 | 8882 | 8862 |
| Trigonidiinae TrigoSP04 | s7n050a | 41 | 45.6 | 45.6 | 8428 | 8212 |
| Trigonidiinae TrigoSP05 | s6n063 | 57 | 47.9 | 41.2 | 8774 | 8774 |
| Trigonidiinae TrigoSP05 | s6n056 | 86 | 52.3 | 44.5 | 8318 | 8290 |
| Trigonidiinae TrigoSP06 | s7n058 | 60 | 26.6 | 26.1 | 7340 | 7308 |
| Trigonidiinae TrigoSP06 | s6n005 | 66 | 23.5 | 22.6 | 7952 | 7876 |
| Trigonidiinae TrigoSP06 | s7n060 | 69 | 24.2 | 23.5 | 7702 | 7640 |
| Trigonidiinae TrigoSP06 | s7n079w | 62 | 25.3 | 24.1 | 7295 | 7268 |

then compared with the available feature database. Using graphical visualisation, a selection of 'similar' songs can be made, searching for nearest neighbors along the frequency and pulse-distance axes. The resulting short list of similar songs from the database can then be compared with the 'new' song, using traditional sonograms and acoustical comparison by the human observer, which includes secondary song features. In the future, neural network approaches could be used for a complete automatization of song classification.

For the cricket fauna from one locality of lowland Amazonian rainforest, Nischk (1999) could show that clusters of acoustic parameters coincide with morphospecies, which are well-documented by voucher specimens. However, it is highly probable that similar songs at distinct sites are produced by distinct species, resulting in "acoustical vicariance". Our automatized feature extraction will facilitate the detection of such "acoustic equivalent species", as well as mapping of ranges and contact zones.

Are "acoustic communities" saturated within a certain parameter space, or are they convergent if we compare distinct faunas? These and similar questions can only be answered by comparative studies, using major databases including voucher specimens, songs and a simple set of sound parameters. Such multimedia databases need not necessarily be centralized, because modern web technologies provide efficient protocols for access to federated databases (see GBIF, www.gbif.org). For interoperability, certain descriptors will be necessary for efficient characterisation. Though traditional song analysis will always be useful, it seems unrealistic that all hitherto digitized sound files could be annotated without using automatic feature extraction software, such as the program presented here.

The present study brings us back to the very first formalized approaches to annotate insect songs, which were the musical an-

notations by Yersin (*cf.* Ragge & Reynolds 1998). To administer multimedia data such as sound files in a database, we need simple qualitative descriptors, such as the numeric values for carrier frequency, pulse rates *etc.* However, this works only for the highly stereotyped songs of Grylloidea. In the case of more complex or broad-band songs, such as those produced by many katydids (Tetragonioidae) and grasshoppers (Acridoidea), other descriptors will be needed. These might consist of simplistic descriptors of power spectra, such as frequency maxima and broadness, as given by the Q-value. The development of innovative descriptors, their simple automated calculation, and tools for annotation and retrieval within large databases is the challenge for the future.

Acknowledgements

This work has been partially supported by the DFG (Deutsche Forschungsgemeinschaft) under SCHW 623/4-2, and a DFG travel grant to K. Riede (Ri 1525/2-1 KON1275/2005).

References

- Desutter-Grandcolas L., Nischk F. 2000. Chant et appareil stridulatoire de deux Trigonidiinae originaires d'Équateur (Orthoptera: Grylloidea: Trigonidiidae). *Annales de la Société entomologique de France* (N.S.) 36: 95-105.
- Dietrich C., Palm G., Riede K., Schwenker F. 2004. Classification of bioacoustic time series based on the combination of global and local decisions. *Pattern Recognition* 37: 2293-2305.
- Faber A. 1953. Laut- und Gebärdensprache bei Insekten. *Orthoptera*

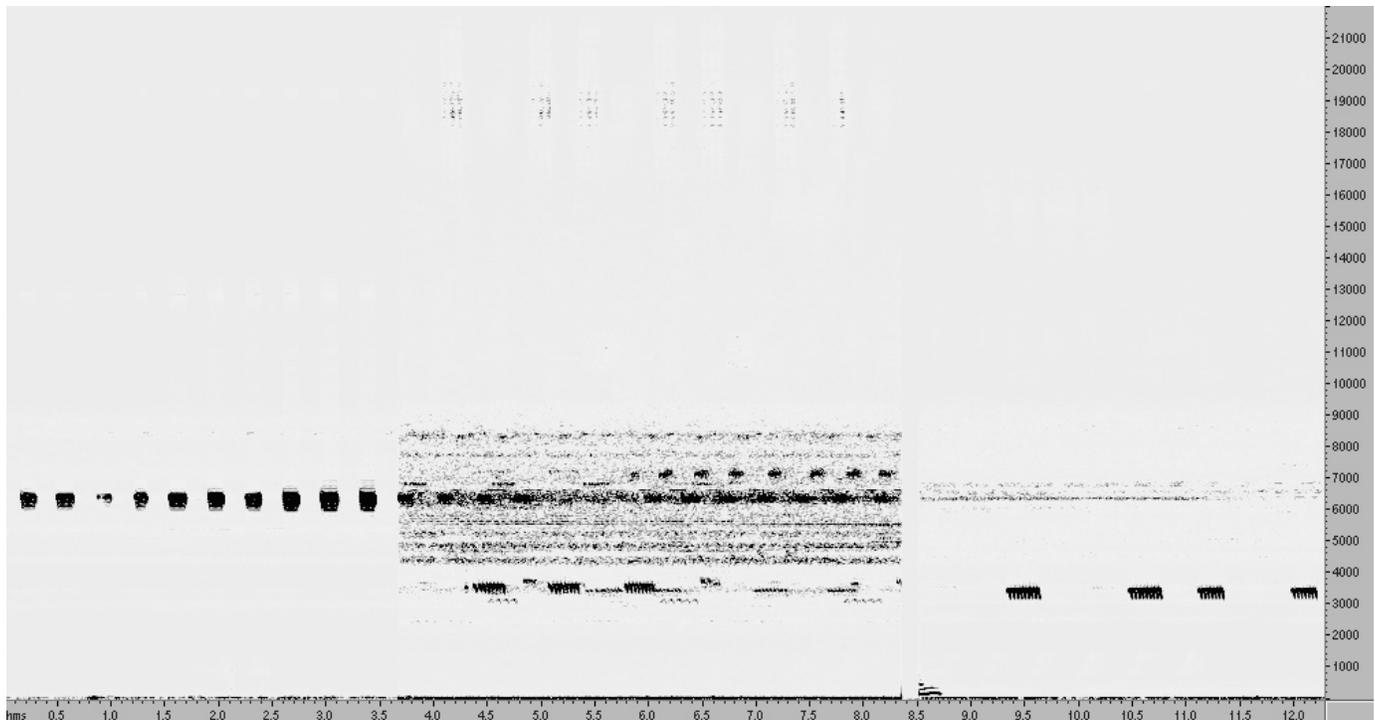


Fig. 6. Sonograms of mixed recording and separated components. Middle block, hms 3.6-8.3: several species sing simultaneously in this recording. It was stored as *Aclodes chamacoru* (s6n033.wav) in the SYNTAX database. The TOSRA software annotated features corresponding to EneopSP2, which is also visible in the mixed recording. For comparison, clear recordings for each species are shown to the right (*Aclodes chamacoru* - s6n025.wav, hms 8.4-12.2) and the left (EneopSP2, based on s6n070c.wav, hms 0-3.5). Note that TOSRA parameter extraction was correct for all clear recordings.

(Geradflügler). Gesellschaft der Freunde und Mitarbeiter des Staatlichen Museums für Naturkunde, Stuttgart.

Helversen, D. von, Helversen, O. von 1998. Acoustic pattern recognition and orientation in orthopteran insects: parallel or serial processing. *Biological Cybernetics* 79: 467-476.

Huber F., Moore, T.E., Loher, W. 1989. *Cricket Behavior and Neurobiology*. Cornell University Press, Ithaca & London.

Ingrisch S., Riede K., Lampe K.-H., Dietrich C. 2004. DORSA - A "Virtual Museum" of German Orthoptera Collections. *Memorie della Società Entomologica Italiana* 82 [2003]: 349-356.

Nischk F. 1999. Die Grillengesellschaften zweier neotropischer Waldökosysteme in Ecuador. Dissertation, University of Cologne.

Nischk F., Otte D. 2000. Bioacoustics, ecology and systematics of Ecuadorian rainforest crickets (Orthoptera: Gryllidae: Phalangopsinae), with a description of four new genera and ten new species. *Journal of Orthoptera Research* 9: 229-254.

Otte D. 1994. *The Crickets of Hawaii*, 396 pp. The Orthopterists' Society, Philadelphia.

Ragge D.R., Reynolds W.J. 1998. *The Songs of the Grasshoppers and Crickets of Western Europe*. Harley Books, Colchester.

Riede K. 1993. Monitoring biodiversity: analysis of Amazonian rainforest sounds. *Ambio* 22: 546-548.

Römer H., Krusch M. 2000. A gain control mechanism for processing of chorus sounds in the afferent auditory pathway of the bushcricket *Tettigonia viridissima* (Orthoptera: Tettigoniidae). *Journal of Comparative Physiology A* 186: 181-191.

Schildberger K. 1994. The auditory pathway of crickets: adaptations for intraspecific acoustic communication, pp. 209-225. In: Schildberger K., Elsner N. (Eds) *Neural Basis of Behavioural Adaptations*. Gustav Fischer Verlag, Stuttgart.

Schwenker F., Dietrich C., Kestler H.A., Riede R., Palm G. 2003. Radial basis

function neural networks and temporal fusion for the classification of bio acoustic time series. *Neurocomputing* 51: 265-275.

Schwenker F., Kestler H.A., Palm G. 2001. Three phase learning for radial basis function networks. *Neural Networks* 14:439-458.

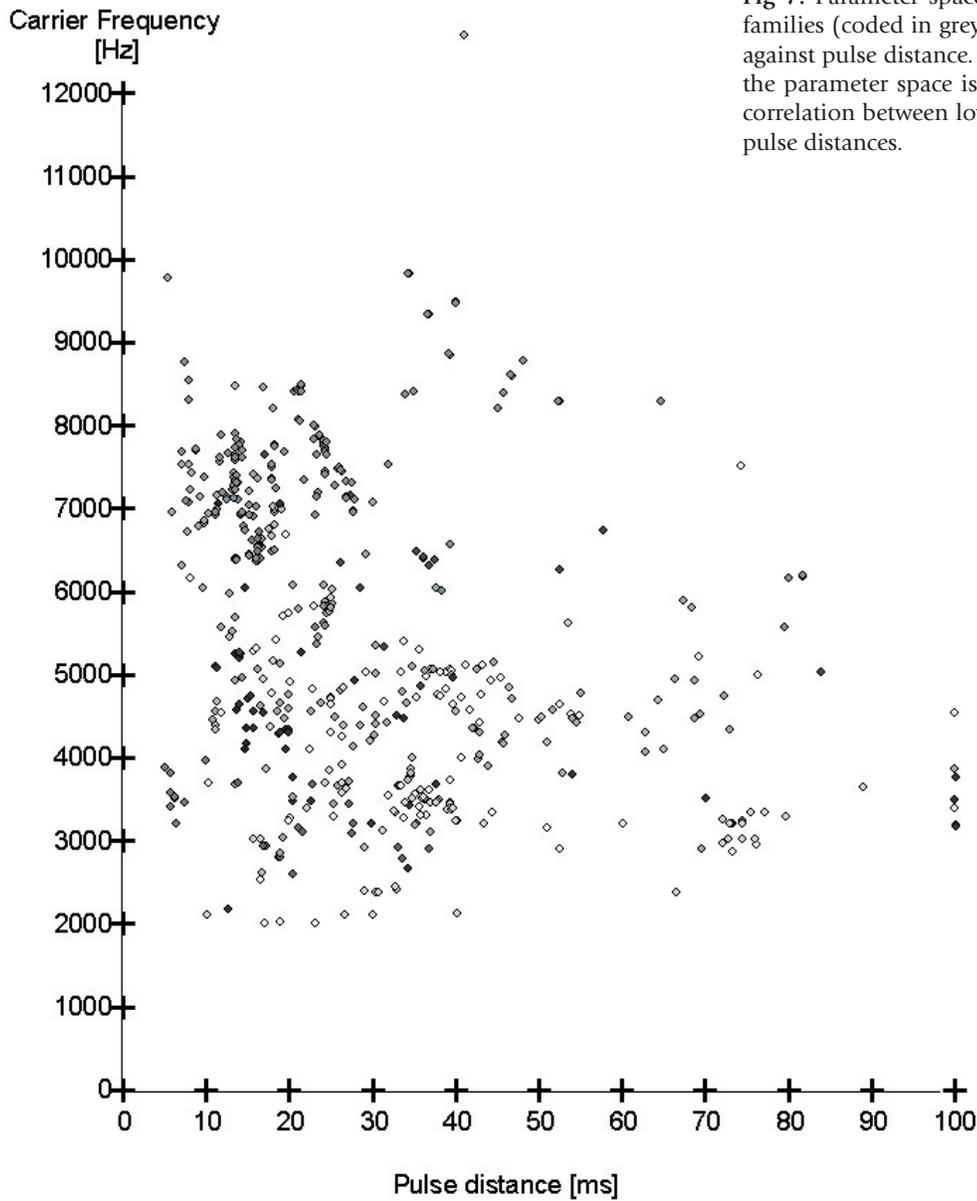


Fig 7. Parameter space of all 625 cricket species and 10 families (coded in grey). Song carrier frequency is plotted against pulse distance. Note that the upper right corner of the parameter space is comparatively empty, indicating a correlation between low song carrier frequency and larger pulse distances.