Listening in the Wild

One-day research workshop · June 25th 2013 · Queen Mary University of London

9:45 Registration, tea and coffee

10:30 Welcome (Mark Plumbley and Dan Stowell)

10:45 Session 1 (Chair: Alan McElligott)

* Thierry Aubin (Université Paris Sud, Orsay)
  Communication in seabird colonies: vocal recognition in a noisy world

* Marc Naguib (Netherlands Institute of Ecology, Wageningen)
  Noise effects on communication in song birds

* Maria Chait (University College London)
  Change detection in complex acoustic scenes

* Jon Barker (University of Sheffield)
  Machine listening in unpredictable "multisource" environments:
  Lessons learnt from the CHiME speech recognition challenges

12:45 Lunch and poster session

14:00 Session 2 (Chair: Dan Stowell)

* David Clayton (Queen Mary University of London)
  Investigating a link between vocal learning and rhythm perception
  using the zebra finch as a model animal

* Rachele Malavasi (Institute for Coastal Marine Environment, Oristano)
  Auditory objects in a complex acoustic environment: the case of bird choruses

15:00 Tea and coffee

15:30 Session 3 (Chair: Mark Plumbley)

* Richard Turner (University of Cambridge)
  Auditory scene analysis and the statistics of natural sounds

* Dan Stowell (Queen Mary University of London)
  Machine listening for birds: analysis techniques matched to the
  characteristics of bird vocalisations

* Mathieu Lagrange (IRCAM, Paris & IRCCYN, Nantes)
  Machine Listening in Complex Environments: Some challenges in
  understanding musical and environmental sounds

17:00 Close, opportunity to continue discussions in a nearby pub/restaurant

Listening in the Wild: speaker abstracts

Communication in seabird colonies: vocal recognition in a noisy world

Thierry Aubin (Université Paris Sud, Orsay)

Most animals rely on acoustic communication for species or individual recognition mate selection and territorial defence, but degradation of sound waves during transmission and irrelevant background noise often limit the ability of receivers to detect and discriminate the signal. A growing literature now focuses on the problem of environmental noise for acoustic communication but the studies concern almost exclusively anthropogenic noises or biotic (competitive signals emitted by other species) and abiotic (wind, rain, torrent) signals. Interfering noise makes signals difficult to detect, and information encoded within the signals becomes harder to extract. Species overcome these interferences either by increasing the amplitude of their vocalisations (Lombard effect) or by shifting up or down the bandwidth frequency of their signals or by using temporal avoidance strategies. Nevertheless, the most challenging conditions for communication by sounds in the noise occur in chorusing frogs and bird colonies. Indeed, the only acoustic signals which cover exactly in the same time the same frequency band and in theory would lead to a perfect masking effect are the vocalisations emitted by conspecifics. In these conditions of competing noise, the acoustic properties of the masker and those of the signaller are similar. The jamming effect is important from an amplitude, time, and frequency point of view, and this increases the difficulty for a receiver to extract the information provided by the sender. On the basis of results obtained in numerous field studies focusing on seabirds, we show that these species use a particularly efficient “anti-confusion” and “anti-noise” acoustic coding system, allowing an accurate identification and localization of individuals on the move in a noisy crowd.The study of these biological models allows us to highlight the basic rules that govern the identification of a precise acoustic message in the noise.

Noise effects on communication in song birds

Marc Naguib (Behavioural Ecology Group, Animal Science Department, Wageningen University, The Netherlands)

Acoustic signals, such as bird song, are often used over long distances and/or in noisy areas, so that information in signals will be degraded or masked. Moreover, noise can indirectly affect communication by acting as general stressor or by shifting attention away from relevant signals. Yet, any such effects of noise depend on both, the nature of the noise and the characteristics of the individuals exposed to it. Here, I will outline some basic constraints of communicating in the wild by focusing on adaptations of signal structures and receiver behaviour to long range communication. I will also present some recent experiments in which we investigated the influence of noise differing in spectral characteristics on (a) male song and breeding success and (b) on parental nest box visits and nestling begging in field population personality-typed wild great tits (Parus major). The results indicate that effects of noise depend on the nature of noise as well as on characteristics of the exposed individuals, providing new insights in how disturbance can differently affect individuals in a population.

Change detection in complex acoustic scenes

Maria Chait (University College London)

The ability to detect sudden changes in the environment is critical for survival. Hearing is hypothesized to play a major role in this process by serving as an 'early warning device', rapidly directing attention to new events. Here we investigate listeners' sensitivity to changes in complex acoustic scenes - what makes certain events 'pop-out' and grab attention while others remain un-noticed? We use artificial 'scenes' populated by multiple pure-tone components, each with a unique frequency and amplitude modulation rate. Importantly, these scenes lack semantic attributes, which may have confounded previous studies, thus allowing us to probe low-level processes involved in auditory change perception. Our results reveal a striking difference between 'appear' and 'disappear' events. Listeners are remarkably tuned to object appearance: change detection and identification performance are at ceiling, response times are short, with little effect of scene-size, suggesting a

pop-out process. In contrast, listeners have difficulty detecting disappearing objects, even in small scenes: Performance rapidly deteriorates with growing scene-size, response times are slow, and even when change is detected, the changed component is rarely successfully identified. I will also report on a series of experiment where we introduced irrelevant interruptions to the scenes such as gaps and clicks and experiments in which we manipulated the patterning of scene components.

Machine listening in unpredictable "multisource" environments: Lessons learnt from the CHiME speech recognition challenges

Jon Barker (University of Sheffield)

Most existing speech technologies operates under the assumption that there will either be very little background noise or that the background is easily modelled (e.g. dominated by a highly predictable component). Research has been driven by mostly unrealistic evaluations that have promoted simple approaches that then fail to operate successfully in real environments. Progress requires evaluations that better reflect the challenge of everyday listening, chiefly, dealing with multiple, unpredictable sound sources. As a step in this direction we recently launched "the CHiME challenge for speech separation and recognition in multisource environments" -- an initiative to bring together the speech technology, signal processing and machine learning communities in a bid to stimulate fresh multidisciplinary approaches to environmental noise robustness. The CHiME challenge employs binaural audio recorded in a family home. The challenge captures both the **difficulties** posed by a noise background composed of contributions from multiple sound sources but also the **opportunities** that arise from having large amounts of continuously recorded data from which to learn. This talk will provide an overview of the CHiME challenge -- now in its second iteration – looking at the evolution of the speech recognition task and at the design of systems that have performed well. While looking at these systems we will consider the general question, "How can prior experience of the acoustic environment be effectively and efficiently exploited in machine listening systems?"

Investigating a link between vocal learning and rhythm perception using the zebra finch as a model animal

David Clayton (Queen Mary University of London)

Though almost all humans tap to a beat spontaneously and with ease, no evidence suggests this extends to our nearest primate relatives. Data from a range of species have led to a working hypothesis that vocal learning may be a prerequisite for beat perception. In common with humans, zebra finches have vocal learning, a critical period for learning song, and similar basal ganglia and cerebellar circuitry. We aim to formalise the zebra finch as an animal model of rhythm perception in order to investigate the dissociation between beat-based timing, which is basal-ganglia dependent, and interval-based timing, which is cerebellar-dependent. The pharmacological and behavioural methodologies that can be conducted using zebra finches will provide an opportunity to study the development of the link between vocal learning and rhythm perception.

Auditory objects in a complex acoustic environment: the case of bird choruses

Rachele Malavasi (Institute for Coastal Marine Environment, Oristano)

The perceptual grouping of spectrotemporal acoustic patterns seems to be highly influenced by the temporal coherence of sound elements. Starting from the temporal feature, others (frequency, pitch) are then collected and linked to the same sound source. At the very basis, this seems to be the most reliable hypothesis explaining how an individual distinguishes a sound source among several others, like a frog call in a chorus. But for birds the picture is particularly complex. Groups of conspecifics may perform together in what is called a group-specific vocal signature, or even birds of different species may sing at the same time and produce an emergent signal. In these cases, it is the attention of the listener that initiates the binding process, so that temporally coherent sounds may be perceived as part of the same auditory object, even if they come from different physical sources.

Auditory scene analysis and the statistics of natural sounds

Richard Turner (University of Cambridge)

It has long been suspected that the auditory system listens attentively to the way in which the energy in each frequency band of the input signal varies over-time. This aspect of sounds is called the modulation content. Recent psychophysical and electrophysiological work has confirmed that modulation appears to be a key quantity.

In this work we show that simple synthetic sounds with different patterns of modulation are indistinguishable from familiar audio-textures including running water, crackling fire, howling wind, snapping twigs, and pattering rain. In fact, we provide methods for learning the modulation patterns from a training sound, which are then used to produce novel realistic-sounding synthetic versions.

We hypothesise that more complicated sounds can be produced by considering richer, hierarchical descriptions of the modulation. This perspective suggests that the modulation patterns are primitive features which might form a fundamental substrate for auditory perception. We show that such a hypothesis, when wedded to the idea that the brain is performing Bayesian inference, is consistent with a large number of psychophysical phenomena, including the Gestalt grouping rules, comodulation masking release, and the continuity illusion. In other words, primitive auditory scene analysis might be thought of as statistical inference for the modulation content of a sound.

Machine listening for birds: analysis techniques matched to the characteristics of bird vocalisations

Dan Stowell (Queen Mary University of London)

Researchers applying machine listening to bird vocalisations often use techniques adapted from automatic classification or speech recognition. However, these techniques imply some assumptions about the sound that can be strongly violated in bird vocalisations. Particular problems include: there may be many birds vocalising in a given audio recording; many birds emit structured patterns of sound and silence rather than continuous sound; and bird vocalisations often contain rapid frequency modulations, which are obscured by spectrogram or MFCC representations. We will describe two aspects of our work which aim to adapt machine listening to accommodate these characteristics. First, we introduce a *Markov renewal process* model for tracking multiple intermittent sources. Second, we consider signal processing that goes beyond the spectrogram to capture fine modulation details. We demonstrate that these adaptations lead to improved performance in tasks such as multiple bird tracking and species recognition.

Machine Listening in Complex Environments: Some challenges in understanding musical and environmental sounds

Mathieu Lagrange (IRCAM, Paris & IRCCYN, Nantes)

Designing computational auditory models raises questions about how humans and animals actually make sense of a complex sound environment and how those process can be computationally implemented. In this talk, we will cover some aspects of this problem by discussing recent results that provide answers to the following questions:

- Is there a need for segregating elements of interest?
- What models can be relevant for representing complex scenes in a generic way?
- Is it meaningful to evaluate computational systems using artificial data?

Examples will be provided using experiments done with musical and environmental audio corpora.

Listening in the Wild: poster abstracts

Social and genetic influences on goat kid calls during development

EF Briefer[1,2] and AG McElligott[1]

1 Biological and Experimental Psychology, Queen Mary University of London, School of Biological and Chemical Sciences, London, UK. 2 Institute of Agricultural Sciences, ETH Zürich, Zürich, Switzerland.

Vocal signals of kinship and group membership are crucial in social species. They allow conspecific recognition and hence facilitate social interactions and cohesion between group members, or between parents and offspring. Most mammals (including goats) are believed not learn their vocalisations, and therefore social experience is believed to be unimportant during vocal development. Similarities between individuals are assumed to arise from inherited physical characteristics of the vocal apparatus. We investigated if and when signals of kinship and non-inherited signals of group membership arise, during vocal ontogeny in goat kids. Four groups of goat kids with the same father and born at different times of year, were recorded at one week and five weeks postpartum. We assessed potential vocal correlates to kinship and group membership. We found that calls of twins were more similar than calls of half-siblings at five weeks but not at one week postpartum. Kinship influenced some vocal parameters at one week, and most parameters at five weeks. Surprisingly, calls of half-sibling kids were more similar when they had been raised in the same group than in different groups, at both one week and five weeks. Our results suggest that goat kid vocalisations show some plasticity at an early age and that social environment shapes vocal development. A potential function for this plasticity is an increase in call similarity between conspecifics that could facilitate kin and group recognition (e.g. mother-offspring recognition). Such findings are unexpected in an ungulate with a relatively simple vocal communication system. Our results support the growing and controversial evidence that mammal vocalisations are more plastic than previously believed.

Artificial Neural Network approach to assess vocal identity, kinship and ageing in goats (Capra hircus)

L Favaro[1], EF Briefer[2] & AG McElligott[3]

1 Department of Life Sciences and Systems Biology, University of Torino, Via Accademia Albertina 13, 10123 Turin, Italy
2 Institute of Agricultural Sciences, ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland
3 Biological and Experimental Psychology, Queen Mary University of London, School of Biological and Chemical Sciences, Mile End Road, London E1 4NS, UK

Machine learning techniques are becoming a more important tool in studying animal vocal communication. The goat (Capra hircus), is a very social species where vocal communication and recognition, particularly between mothers and offspring, is important. In this study we tested the reliability of a Multi-Layer Perceptron (feed-forward ANN) to automate the process of vocal analysis for individuality, group membership and ageing in this species. Vocalisations were obtained from 10 half-sibling (same father but different mothers) goat kids, belonging to 3 distinct social groups. We recorded 164 contact calls emitted during early postnatal days, and 157 additional calls recorded from the same individuals at 5 weeks. For each call we measured 27 spectral and temporal acoustic parameters using automatized procedures in PRAAT (www.fon.hum.uva.nl/praat). For each classification task we built stratified 10-fold cross-validated neural networks using the WEKA software package (www.cs.waikato.ac.nz/ml/weka). The input nodes corresponded to the acoustic parameters measured on each signal. ANNs were trained with the error-back-propagation algorithm. The number of hidden units was set to the number of attributes + classes. Each model was trained for 300 epochs (learning rate 0.2; momentum 0.2). To estimate a reliable error of the models, we repeated 10-fold cross-validation iterations 10 times and calculated the average predictive performance. The Correctly Classified Instances was 71.13±1.16% for the vocal individuality, 79.59±0.75% for the social group and 91.37±0.76% for the age of the vocaliser. Our results demonstrate that ANNs are a powerful tool for studying contact calls. The performances we achieved are higher than those previously obtained with classical statistical methods such as Discriminant Function Analysis. Further development of this approach might include the classification of contact calls within other social species and the comparison of ANNs with other machine learning techniques.

Listening to the fish – Acoustic monitoring of freshwater biodiversity through audio signature recognition

Simon Linke [1], Toby Gifford [2], Mark Kennard [1]
1 Australian Rivers Institute, Griffith University, Brisbane, Australia
2 Queensland Conservatorium, Griffith University, Brisbane, Australia

Accurate monitoring of freshwater species distributions and abundances is critical for effective management and conservation of freshwater ecosystems, as well as for species-specific conservation measures. Freshwater fish are widely used as a key indicator organism but accurately quantifying their population status is often achieved by using invasive techniques such as electrofishing or netting. Apart from logistical and ethical challenges, the active nature of the survey techniques can also bias the result. Aquatic audio monitoring to date has concentrated on sonar echo profiles, however recent advances in bioacoustics and machine learning make monitoring, detection and identification of individual fish sounds feasible. We are proposing a non-invasive real-time technique that records acoustic communication signals from fish using a hydrophone, and applies digital audio signature recognition techniques to identify the species, as well as estimating relative abundance. The methodology will be tested on two main target groups of species in Australia's north: the Terapontid grunters (e.g. Sooty grunter - Hephaestus fuliginosus) as well Neoariid catfishes (e.g. fork-tailed catfish - Arius graeffei) - but could be extended to North American or South East Asian species through a collaboration with international colleagues.

Auditory onsets and salience

A. Kovács[1], M. Coath[2], T. Bőhm[1], S. Denham[2], I Winkler[1]
1 Institute of Cognitive Neuroscience and Psychology, Hungarian Academy of Science, Budapest, Hungary
2 Cognition Institute, Plymouth University, Plymouth, UK

Enhancement of auditory transients is well documented in the auditory periphery and mid-brain and it is also known that transients are important in, for example, speech comprehension, object recognition and grouping. In this work we introduce the novel approach of using an artificial neural network to implement a model of auditory transient extraction which is based on the assymetry of the distribution of energy inside a frequency dependent time window. We compare the output using this method with the original model, and with other methods of identifying salient events in an auditory stimulus motivated by phonological (Landmarks) and information theoretic (Bayesian Surprise) analysis of human speech. The results show that the transients identify, in most cases, the same stimulus features as the comparison methods and, for comparison, we show preliminary results from similar analysis of other behaviourly relevant sounds (eg bird song) and from mixtures of sounds.

Acoustic Classification of Goose Behaviours

Kim Steen
School of Engineering, Aarhus University

Throughout the world, damage caused by wildlife creates significant economic challenges to human communities. Since human-wildlife conflicts are increasing the development of cost-effective methods for reducing damage or conflict levels is important in wildlife management. Here we present a new concept, circumventing the problems of habituation, which is often the major limitation on scaring devices, for the species-specific detection of an animal species causing conflict with agriculture.

Variation in machine-listening performance from long-term kiwi monitoring

Andrew Digby[1], Michael Towsey[2], Ben Bell[1], Paul Teal[3]
1: School of Biological Sciences, Victoria University of Wellington, New Zealand
2: Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
3: School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

The New Zealand kiwi are threatened and iconic species that are mostly monitored by acoustic methods. Existing surveys rely on field call counts, or at best manual inspection of spectrograms, so the use of autonomous acoustic recorders is restricted by data processing constraints. To address this we have developed a kiwi call detector using discrete cosine transform and decision tree methods. Verification from continuous 3-

year recordings of little spotted kiwi (Apteryx owenii) in their natural habitat shows that the detector performs well and can greatly improve call survey efficiency. This large, long-term dataset also reveals significant temporal variation in detector performance, as a result of changes in `noise' from other species. This highlights the importance of considering the full range of acoustic environments encountered in field conditions when assessing the performance of machine-listening methods.

Bioacoustics recordings as a memory of the natural world and the Portuguese Natural Soundscape Project

Marques, Paulo A. M.
UIEE, ISPA-Instituto Universitário, and Museu Nacional de História Natural da Ciência, Universidade de Lisboa, Portugal

Sound recording are primary sources of information that capture the acoustic environment of a place in a time. If preserved these recordings can safeguard the “acoustic memory” for use by future generations. The more recordings the better will be the acoustic model of the natural world and it value to its understanding. This acoustic model if broadly accessible can be used in very different context taking advantage of its historical, aesthetical or scientific value. Preserved recordings can be re-visited to verify research and re-used to test new hypothesis derived form technological and conceptual development. The Portuguese Natural Soundscapes Project aims to portrait the contemporaneous Portuguese natural soundscapes, to build an acoustic memory as a legacy for future generations. The recordings document some of the best-preserved locations and representative soundscapes. We are sampling each soundscape with 24 hours continuous recording using a 5 omnidirectional microphones array over a 35m circle and other smaller duration recordings including focal recordings with directional microphones. With this approach we document biodiversity but also register the human presence. The project sampled 21 locations distributed by Portugal with 2700 hours of recorded material. These recordings are being processed and are deposited in the Portuguese Natural sound Archive for safekeeping.

A Comparison of Non-stationary Methods and Models for birdsong analysis

Sašo Muševič
Music Technology Group, Universitat Pompeu Fabra , Barcelona, Spain

Bird chirps are known to exhibit extremely high frequency modulations (up to 100kHz/s) as well as significant amplitude modulation simultaneously. The combined AM/FM effect reduces the readability of the spectrogram and this negatively impacts any machine learning or other high-level algorithm that relies on it. Non-stationary time-frequency reallocation methods can drastically compact such blurred representations and increase readability, beneficial for improved visualization as well as accurate sinusoidal parameter estimation.

Discriminate an auditory “figure” from ground – an MEG study

Teki, S.[1,2], Payne, C.[2], Griffiths, T.D.[1,3], Chait, M[2]
1. Wellcome Trust Centre for Neuroimaging, University College London, UK. 2. UCL Ear Institute, University College London, UK. 3. Auditory Group, Institute of Neuroscience, Newcastle University, UK. (Equal contributions for Griffiths and Chait)

The natural acoustic environment comprises of a complex mixture of sounds. In order to isolate individual sounds of interest, the mixture needs to be parsed. This is a highly complex task whose underlying brain mechanisms are still not understood. In order to model naturalistic acoustic scenes, we developed a stochastic figure-ground stimulus (SFG, Teki, Chait et al., 2011). The stimulus comprises a series of chords (25 ms long) containing random frequencies that vary from one chord to another. To study segregation, we introduced a figure by randomly selecting a certain number of frequencies ("coherence") and repeating them over a certain number of chords. This allowed us to control the salience of the figure, which can only be extracted by binding across time and frequency. We found that listeners are very sensitive to the emergence of these complex figures. We previously established a role for the intraparietal sulcus (IPS) in stimulus-driven segregation of these figures (Teki, Chait et al., 2011) and modeling work suggests a role for temporal coherence in mediating segregation (Shamma et al., 2011). To understand the brain dynamics of segregation, we used MEG. Figures

with different salience (coherence of 2, 4 or 8; 0.6s long) were presented after statistically similar background segments (0.6 s). Listeners were engaged in a visual task and were naive to the SFG stimulus. In another condition, we presented the same stimuli but interspersed with alternating white noise, as we previously found that this does not affect figure detection (Teki et al., 2012). Analysis of time-locked activity in the auditory cortex shows an initial onset response to the emergence of the figure, followed by a sustained response which follows the figure. The figure onset responses occur about 100 ms in the coherence=8 condition, and 150 ms for coherence values of 4 and 2. These latencies correspond to duration of 4 (or 6) chords and parallel behavioral performance (obtained separately). Latencies from the 'noise' condition reveal the same threshold, suggesting that the segregation mechanism was not affected by the interspersed noise bursts.

Source analysis based on fitting of equivalent current dipoles suggests that a model comprising four sources in bilateral planum temporale (PT) and IPS explains the post-transition responses to the figure, during both the early and sustained phases. Results from ongoing analyses of effective connectivity between the temporal and parietal sources as well as time-frequency analysis will be presented.

Experimental evidence for signals of quality and motivation in fallow deer (Dama dama)

Benjamin J. Pitcher[1], Elodie F. Briefer[2], Elisabetta Vannoni[3] and Alan G. McElligott[1]
1. Queen Mary University of London, Biological and Experimental Psychology, School of Biological and Chemical Sciences
2. Animal Behaviour, Health and Welfare Unit, Institute of Agricultural Sciences, ETH Zürich
3. Department of Functional Neuroanatomy, Institute of Anatomy, University of Zurich

Vocalisations encode a range of information about the caller which may or may not be intentionally produced. Fallow deer bucks only vocalize ('groan') during the breeding season. Males groan up to 3000 groans times per hour, and lose about 25% of their body weight during the rut. Groans are individually distinctive and encode the quality of the caller, reflecting changes in dominance between years. Males modulate their calling rates, calling faster when both other males and females are nearby. Groans also reveal caller fatigue; becoming shorter and higher pitched towards the end of the rut. However, no studies have investigated how information in groans is perceived and used. Using playback experiments, we investigated the roles of groaning rate and caller fatigue on the perception of the threat posed by the caller. Males significantly increased vigilance to higher than lower calling rates. Furthermore, males attended more to early rut calls, than to late rut calls containing fatigue-related effects. Our results demonstrate that fallow bucks can perceive both changes in calling rate and fatigue induced effects in groans. Fallow bucks extract honest information from groans about the contemporary quality of the caller, simultaneously gaining information about caller motivation and fatigue.

Monitoring temporal change of bird communities with dissimilarity acoustic indices

Laurent Lellouch[1,4], Sandrine Pavoine[2,3], Frédéric Jiguet[2], Hervé Glotin[4], Jérôme Sueur[1],*
1 Muséum national d'Histoire naturelle, Département Systématique et Évolution, Paris, France.
2 Muséum national d'Histoire naturelle, Département Ecologie et Gestion de la Biodiversité, Paris, France.
3 Mathematical Ecology Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford, UK.
4 Université Sud Toulon Var, UMR CNRS 7296 LSIS, France.

An avian community can be characterized by its species composition as well as by the soundscape it produces. The development of dissimilarity acoustic indices aims at describing changes in the composition of animal communities, relying on recorded data. Recordings of dawn chorus from March 24th to June 5th 2009 at three different sites in Haute-Vallée de Chevreuse (France) were analysed by aural identification, making possible a realistic simulation of acoustic communities. We estimated whether five spectral dissimilarities (correlation-based, Kullback-Leibler, Kolmogorov- Smirnov, pointwise difference, cumulative frequency dissimilarity) applied on field recordings and simulated communities could provide a good description of the compositional changes that affect avian communities. This was tested at two different temporal scales: a) do acoustic indices give a good description of the global turnover of species during a whole season? b) do acoustic indices enable to detect dates at which the acoustic community composition is rapidly changing?

Listening to the Environment: Hearing Differences from an Epigenetic Effect in Solitary and Gregarious Locusts

Shira D. Gordon[1]*, Joseph C. Jackson[1], Steve M. Rogers[2], James F.C. Windmill[1]

1 Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow
2 Department of Zoology, University of Cambridge—current address School of Biological Sciences, University of Sydney

Developmental plasticity enables animals to remain fit as a result of which micro-environment they experience during growth. Locusts exhibit epigenetic effects resulting in two phases (solitary and gregarious) differing in appearance, behaviour, and visual capabilities. Recently, it has been shown that gregarious animals have vision that is more specialized for living in a swarm. Conversely, for hearing, our results show solitary locusts are more sensitive, presumably to hear their predators more precisely as they are not protected by the numbers of the swarm. Neural sensitivity experiments signify a better ability to resolve decibel levels (low frequencies) and a greater neurological response (80% vs. 65%) (high frequencies). In addition, we link this data with the nanometre mechanical responses of the ear's tympanal membrane to sound, finding that solitary animals exhibit more displacement movement. Finally, we identify significant differences in the shape of the tympana that may be responsible for the hearing sensitivity. In conclusion, the trade-off between enhanced hearing in solitary animals, probably for predator detection, and better vision in gregarious locusts for swarming behaviour, highlights the importance of epigenetic effects set forth during development and begins to identify how animals are equipped to match their immediate environmental needs.

Bioacoustic monitoring in realistic scenarios with an emphasis on periodicity in birdsong

Daniel Wolff[1], Rolf Bardeli[2], Martina Koch[3], Klaus-Henry Tauchert[3], Frank Kurth[4] and Michael Clausen[5]
1 City University, London. 2 Fraunhofer Institute Intelligent Analysis and Information Systems IAIS, Germany.
3 Animal Sound Archive Berlin, Germany. 4 Fraunhofer Institut für Kommunikation, Informationsverarbeitung und Ergonomie (FKIE). 5 Friedrich-Wilhelms Universität Bonn, Institut für Informatik

In 2006, a project was set up aiming at a transfer of established techniques in speech recognition and music information retrieval to problems typically met in bioacoustics. Its focus was set on the automated detection of particular animal species in natural environments. The Animal Sound Archive at Humboldt University, Berlin delivered the precisely annotated data necessary for the development of recognition algorithms. Furthermore, the acquisition of a great amount of acoustic monitoring data was performed by experts from this institution. Afterwards, at the University of Bonn, detection algorithms were developed for specific endangered bird species, utilising newly developed methods for denoising in nature recordings and periodicity analysis. We here sketch the algorithms used for birdsong detection as well as a general overview of the project including special requirements for unsupervised monitoring.

The role of temporal regularity in auditory segregation

Lefkothea-Vasiliki Andreou [1], Makio Kashino [2], Maria Chait [1]
1 UCL Ear Institute, London, UK. 2 NTT Communication Science Laboratories, NTT Corporation, Atsugi, Japan

The idea that predictive modelling and extraction of regularities plays a pivotal role in auditory segregation has recently attracted considerable attention. The present study investigated the effect of one basic form of regularity, rhythmic regularity, on auditory stream segregation. We departed from the classic streaming paradigm and developed a new stimulus, Rand-AB, consisting of two, concurrently presented, temporally uncorrelated, tone sequences (with frequencies A and B). To evaluate segregation, we used an objective measure of the extent to which listeners are able to selectively attend to one of the sequences in the presence of the other. Performance was quantified on a difficult pattern detection task which involves detecting a rarely occurring pattern of amplitude modulation applied to three consecutive A or B tones. In all cases the attended sequence was temporally irregular (with a random inter-tone-interval (ITI) between 100 and 400 ms) and the regularity status of the competing sequence was set to one of four conditions: (1) random ITI between 100 and 400 ms (2) isochronous with ITI = 400 ms. (3) isochronous with ITI = 250 ms (equal to the mean rate of the attended sequence) (4) isochronous with ITI = 100 ms. For a frequency separation of 2 (but not 4) semi tones we observed improved performance in conditions (3) and (4) relative to (1), suggesting that stream segregation is facilitated when the distracter sequence is temporally regular, but that the effect of temporal regularity as a cue for segregation is limited to relatively fast rates and to situations where frequency separation is insufficient for segregation. These findings provide new evidence to support models of streaming that involve segregation based on the formation of predictive models.

An Adaptive Background Model for Real-World Auditory Event Detection

Gineke A. ten Holt, Johannes D. Krijnders, Peter W.J. van Hengel
INCAS, the Netherlands

To process sounds in real-world, noisy environments, distinguishing be-tween foreground and background is an important first step. Since real-world environments are dynamic and changeable, event segmentation cannot be done with a single model for all circumstances. We present an adaptive background model that is capable of incorporating changing background events while leaving foreground events untouched. This model has previously been used in human aggression detection [1] and is adapted for general environmental sound recognition. First, the sound is transformed to a time-frequency representation using a model of the human cochlea [2]. The energy in each frequency channel is attributed to foreground or background in a gradual manner using a sigmoid function. Energy attributed to the background is integrated into the current background using an exponentially decaying function, ensuring that the influence of older inputs on the background decreases. By changing the time constant of the exponential function, the speed with which the background model adapts to new inputs can be adjusted. By combining background models for different adaptation speeds, attention can be focused on auditory events with various onset speeds. One advantage of the model is that foreground status is determined by the difference between background and input energy levels, not by an absolute energy threshold. Thus, the same threshold (sensitivity) can be used for situations wi th different general energy levels (louder or more quiet environments).

Machine Analysis of Bird Vocalisations

Peter Jancovic[1], Munevver Kokuer[1,2], Masoud Zakeri[1], Martin Russell[1]
1 School of Electronic, Electrical & Computer Engineering, University of Birmingham, UK
2 School of Digital Media Technology, Birmingham City University, UK

This poster presents outcomes of our research on machine analysis of bird vocalisations in real-world noisy environments. The work investigates modifications of techniques which have been demonstrated to be effective for automatic speech pattern processing and develop novel techniques that appropriately account for unique properties of bird vocalisations. We first present a method for detection of sinusoidal components in noise. High detection accuracy even in strong noisy conditions is demonstrated. This is employed for detection of bird tonal vocalisations and for providing representation that reflects well the properties of bird tonal vocalisations. As the method provides a decomposition of the entire acoustic scene into sinusoidal components, it can be directly employed for detection of multiple simultaneous bird vocalisations. We then demonstrate the employment of this tonal-based representation for recognition of syllables of bird species in noisy environments. Experimental results show significant recognition accuracy improvements over the use of conventional Mel-frequency cepstral coefficients, in both standard and noise-compensated systems, and a strong robustness to mismatch between the training and testing conditions.

Auditory Cortex is Highly Sensitive to Regularity in Sound Sequences

Nicolas Barascud, Maria Chait
UCL Ear Institute, University College London, UK

We used psychophysics and magnetoencephalography (MEG) functional brain imaging to assess listeners' ability to detect the emergence and violation of complex regularities (characterized by long repeating patterns) in ongoing sound sequences and the degree to which this process is bottom--up driven or dependent on explicit attention. Stimuli were tone pip sequences that contained transitions between random and regular frequency patterns. Transitions from a regularly alternating to a random tone sequence (REG--RAND) are immediately detectable as the first tone to violate the established regularity pattern is sufficient to signal the transition. In contrast, listeners must wait longer (at least one regularity cycle) to detect the opposite transition –from a random to a regular pattern (RAND--REG). We used sequences of 50ms tone pips arranged according to 4 frequency patterns: REG sequences consisted of a regularly repeating pattern of X tones (X=10, 15, 20; new pattern for each trial). RAND sequences consisted of a sequence of tones of random frequencies. REG--RAND and RAND--REG sequences contained a transition between a regular and a random pattern. In all signals, the time of change was jittered across trials. In the behavioural experiments (N=16), subjects were actively detecting the transitions in REG--RAND, RAND--REG (change occurred in 50% of the trials). In the MEG experiment (N=16; different subjects),

naïve participants listened to RAND--REG and REG--RAND stimuli while performing an unrelated visual decoy task. Behavioural response times reveal that subjects required about a cycle and a half to detect the emergence of regularity in RAND--REG signals. Since the transition is not detectable before the first regularity cycle, our results show that listeners only required an additional half cycle to recognize the onset of regularity. These findings are supported by our MEG data, which indicate that brain response latencies (when the subjects were not actively listening to the transitions) reliably match the RT data. Together, our results suggest that the auditory system is remarkably efficient at detecting the appearance and disappearance of regularities in sound sequences, even for very long patterns. Together, our results suggest that the auditory system is remarkably efficient at detecting the appearance and disappearance of regularities in sound sequences, even for very long patterns.

Are there grammars for non-linguistic sounds and what use are they?

Brian Gygi[1], Christian Fullgrabe[2]
1 Nottingham Hearing Biomedical Research Unit. 2 Institute for Hearing Research, Nottingham

The positive effects of top-down linguistic information, such as grammars, in speech are well-known and provide the basis for tests such as the SPIN (SPeech-In-Noise) test (Bilger, Nuetzel et al 1984). Attempts to impose artficial grammars on environmental sounds (Howard and Ballas, 1980) were only partially successful perhaps due to the prexisting semantic content of the sounds used. This work applied a formal grammatical system generated by a Finite State Machine (FSM; Reber, 1976) to 'chimerae'; synthetic sounds created by combining the spectral content of one sound (e.g., a siren) with the envelope of another (such as a bouncing ball). Participants learned the artificial grammar implicitly by being exposed to repeated instances of grammatical sequences of chimerae. After exposure, they were tested on the ability to distinguish grammatical from non-grammatical sequences; participants evinced rapid learning of the grammatical patterns. Next, participants were tested on how well they could generalize the rules of the grammar by having to classify novel patterns of chimerae as grammatical or non-grammatical. The participants overall performed extremely well, suggesting they had learned the rules of the grammar implictly. Finally, the benefit of grammatical sequences was assessed by having participants identify individual chimerae of a sequence while mixed in noise, to simulate difficult listening conditions. The sounds in grammatical sequences were more identifiable than in non-grammatical, replicating the situation found in speech. This suggests that at least in this way speech is not "special;" non-linguistic chimerae are able to be organized in formal grammar that can be learned implictly and show some of the same benefits that grammar provides for understanding speech in difficult listening conditions.

Automatic bird classification based on MFCC clusters, ranked 4th @ ICML4B Kaggle 2013 competition

Olivier Dufour[1,2,4,5], Giraudet Pascale [1,3](a,c), Thierry Artières [5](e), Hervé Glotin [1,2,6]
1. Université du Sud Toulon Var. 2. DYNI team, Information and Systems Sciences Lab LSIS CNRS. 3. Biology department. 4. BIOTOPE. 5. Université Pierre et Marie Curie, Paris 6 Computing Lab LIP6 CNRS. 6. Institut Universtitaire de France

Automatic bird call classification may be an efficient way to monitor endangered population and the health of an ecosystem. At this time, only few species are modelized, and the automatic systems are yielding to poor recall and precision in the case of multiple species. In the framework of the SABIOD ICML for Bioacoustic worskhop (http://sabiod.univ-tln.fr/icml2013) we developed a simple bird classification system for 35 species, based on the baseline Mel Cepstral Coefficients (MFCC) that are used for automatic speech processing. The data for this challenge, copyright of Fernand Deroussen and Jerome Sueur of the Museum National d'Histoire Naturelle, consist in 35 species recordings at high SNR and 90 test files of few minutes each that were recorded in the Parisian area, each morning during few weeks. We first fixed the MFCC parameters according to simple minimal residual criteria to fit as best as possible in average the 35 target species of the train set. Then we cluterized the test set into silence, and two other clusters to modelize different types of call of each species. We finaly train SVM to generate the (90x35) probabilities = P('The species j sings in the test file i'). Our system is ranked 4th on the official benchmark among 77 international teams, with an average AUC of 65%. We then discuss on possible improvements, and integration of external data or recordings (as wikipedia wav samples, or taxonomia for hierarchical classification, ...).

Listening in the Wild

One-day research workshop

June 25th 2013

Queen Mary University of London

Organised by Dan Stowell and Mark Plumbley

dan.stowell@eecs.qmul.ac.uk

mark.plumbley@eecs.qmul.ac.uk

Supported by EPSRC Leadership Fellowship EP/G007144/1