A FRAMEWORK FOR THE ANALYSIS OF MEDICAL DATA USING DATA SONIFICATION ACROSS THE GÉANT NETWORK AND THE EUROPEAN GRID INFRASTRUCTURE

Roberto Barbera
Department of Physics and Astronomy of the University of Catania and Italian National Institute of Nuclear Physics, Division of Catania – Italy
e-mail: roberto.barbera@ct.infn.it
Francesca Falcetta
The Mario Negri Institute for Pharmacological Research, Milan – Italy
e-mail: francesca.falcetta@marionegri.it
Giuseppe La Rocca
Italian National Institute of Nuclear Physics, Division of Catania – Italy
e-mail: giuseppe.larocca@ct.infn.it
Massimo Rizzi
ARCEM – Italian Association for the Research on Brain and Spinal Cord Diseases, Vittuone (MI) and the Mario Negri Institute for Pharmacological Research, Milan – Italy
e-mail: massimo.rizzi@arcem.it
Mariapaola Sorrentino
ASTRA Project and LHC Open Symphony, Cambridge UK
e-mail:mariapaola@astraproject.org
Domenico Vicinanza
DANTE, City House, 126-130 Hills Road, Cambridge CB2 1PQ, UK
e-mail: Domenico.Vicinanza@dante.net

Paper Type
Case study

Abstract
Epilepsy is one of the most common, tough serious, neurological disorders. It has been evaluated that about 3% of people will be diagnosed with epilepsy at some time in their lives. One of the main characteristics of epilepsy is represented by seizures and each type of epilepsy has its own unique combination of seizure type, typical age of onset, EEG findings, treatment, and prognosis.

In this work, “inter-ictal” EEGs (i.e., EEG recording done sufficiently far from a seizure) have been analysed for the first time using a sonification technique and exploiting the European Grid Infrastructure (EGI) in order to identify a baseline condition (hopefully, a marker). The identification of this marker through data sonification may help highlighting and characterising the temporal patterns embedded in the EEGs of epileptic patients, providing a powerful tool for seizure prevention, hence improving the quality of life of people and promoting the research for new therapeutic interventions.

Automated melodic/harmonic analysis algorithms, executed on a set of EGI sites connected by the GÉANT network, have been used.

The submitted contribution reports on the outcomes of a successful multi-disciplinary collaboration among musicians, computer scientists, network experts and neuro-scientists.

Keywords
EGI, GÉANT, Epilepsy, Sonification, Distributed Data Analysis, Science Gateway

1. Sonification as a tool for scientific investigation

Data Sonification is a practical and effective alternative for representing any data set (numbers, strings, text, images, etc.) using sound signals and melodies instead of visual signs and provides data analysis with a different perspective. The human ear is, in fact and to some extent, naturally trained to analyse series of data, detect anomalies and spot (ir-) regularities. It naturally recognises patterns, structures and sequences as a function of time. If one searches a particular

value in a huge background, spotting it on a graph can be difficult, while finding it through a melody can result in a much easier task. Everybody can in fact spot a wrong note in a song, even without any special music training.
Another remarkable possibility associated with data sonification is using well-established, automated techniques to analyse melodies and harmonies to extract information about the structure of the initial data set and its fundamental properties (e.g. symmetries, regularities, etc,).
From a technical point of view, converting numerical information into sound signals is the same as creating a graph from numerical values, except that melody constituents are notes and tones instead of lines and points. In order to have a meaningful sonification process, two conditions must be satisfied:

(i) *Uniqueness*: a single item of data must be linked to one and only one point or sound, and
(ii) *Covariance*: the graph or the melody must vary as quickly as the data. In music, this second condition is satisfied by frequency, timbre and volume.

Data sonification techniques are currently used in several scientific, engineering, and education fields and for many different purposes: surveillance, data analysis, monitoring, etc.
In addition, sonified data provides an effective data analysis and interpretation tool. In medicine, for instance, acoustic data analysis is used every time the physician uses the stethoscope to listen to breath noise and heart tones.

2. **Epilepsy: one of the most common and serious neurological disorders**

In this contribution, we report on the experience of using data sonification in the context of epilepsy studies. Epilepsy is one of the most common, tough serious, neurological disorders. About 3% of people will be diagnosed with epilepsy at some time in their lives[*].
Genetic, congenital, and developmental conditions are mostly associated with it among younger patients; tumours are more likely over age 40 while head traumas and central nervous system infections may occur at any age. The prevalence of active epilepsy is roughly in the range of 5–10 per 1,000 capita.
Epilepsy's approximate annual incidence rate is 40–70 per 100,000 capita in industrialized countries and 100–190 in less developed countries; socio-economically deprived people are at higher risk[†]. About 50 million people worldwide suffer from epilepsy, and nearly 90% of cases occur in developing countries. Epilepsy becomes more common as people age. Onset of new cases occurs most frequently in infants and elder people.

3. **Sonification and epilepsy diagnosis**

One of the main characteristics of epilepsy is represented by seizures and each type of epilepsy has its own unique combination of seizure type, typical age of onset, EEG findings, treatment, and prognosis. In this scientific area, the change of hidden temporal patterns associated to brain electrical activity, as represented by the electro-encephalogram (EEG), may characterise the transition from an apparently normal (or less pathological) brain functioning to a state which may precede the occurrence of a seizure.
Sonification and melody analysis have been taken into account for EEG analysis thanks to their natural time-domain nature (as opposed to frequency-domain analyses possible with techniques based on Fourier transformation).
In this work, “inter-ictal” EEGs (i.e., EEGs recorded sufficiently far from seizures) have been analysed for the first time using a sonification technique and exploiting the European Grid Infrastructure (EGI) in order to identify a baseline condition (hopefully, a marker). The identification of this marker through data sonification may help highlighting and characterising the temporal patterns embedded in the EEGs of epileptic patients, providing a powerful tool for seizure prevention, hence improving the quality of life of people and promoting the research for new therapeutic interventions.
Preliminary results by the application of the recurrence quantification analysis show the existence in the EEGs of epileptic patients of long-range temporal correlations which may be differently expressed depending on the EEG state they belong to. This finding supports and extends the concept of epilepsy as a disease related to the spatio-temporal interactions of complex neuronal networks which may give rise to dynamics expected to be different according to how far the brain is from a seizure-generating critical state.

4. **The role of advanced R&E networking and Grid computing**

Sonification and subsequent data analysis requires enormous amounts of networking and processing power to produce results. Long sequences needs to be generated and analysed using different parameters and different mapping options. High quality audio files and animations (displaying the evolution of the melody) need in fact to be generated for each EEG data sequence.
The whole work described here consequently relied on the pan-European GÉANT network, which operates at a speed of up to 100 Gbps, and on the EGI distributed computing service. Grid computing works by linking together multiple computers at different locations through high speed networks, combining their processing power to deliver faster results

---

[*] Hauser, WA, Kurland, LT (1975). "The epidemiology of epilepsy in Rochester, Minnesota, 1935 through 1967". Epilepsia 16 (1): 1–66
[†] Sander JW (2003). "The epidemiology of epilepsy revisited". Curr Opin Neurol 16 (2): 165–70

when analysing huge volumes of data. Among EGI sites, the most used one in the present analysis have been those belonging to the Italian Grid Infrastructure (IGI).

5. First results

A selection of pictures from the first results prepared by the automated sonification/analysis package and then processed by the neuro-scientists is displayed below. Figure 1 shows the structure of EEG data converted into a melody and displayed in three different ways. Figure 2 displays a frame from an animation created by looking at tonal attractors of the melody coming from the sonification. This kind of animations is particularly useful because they could in future link the evolution of the EEG patterns with the evolution of tonal distribution (i.e. how notes are arranging themselves around a certain focal point). Finally, Figure 3 displays the music interval distribution, highlighting what are the most recurrent data intervals (i.e., the difference between adjacent values).
Concluding, even if it is still in a preliminary phase from a scientific/medical point of view, a quite comprehensive sonification and analysis infrastructure has been put in place, ready to be used by the researchers for their studies on epilepsy.
The next step will be using these outcomes to try to identify "musical markers" able to identify interesting phenomena from the medical point of view.

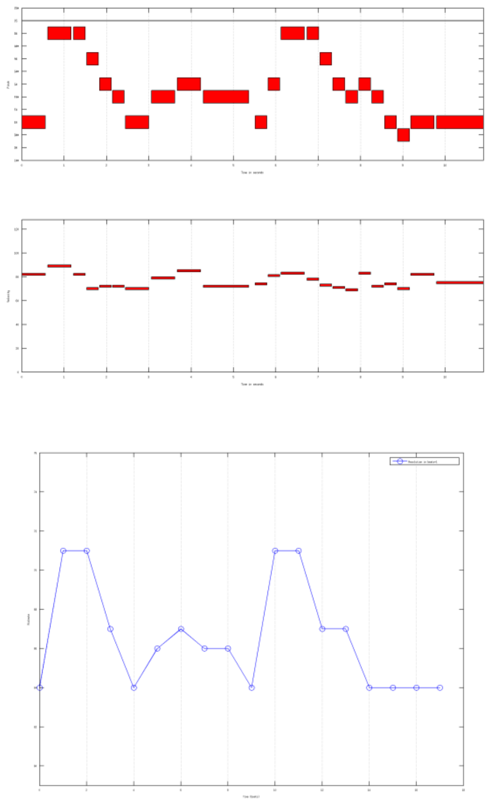

Figure 1- A series of three diagrams showing the melodic structure of an EEG data fragment

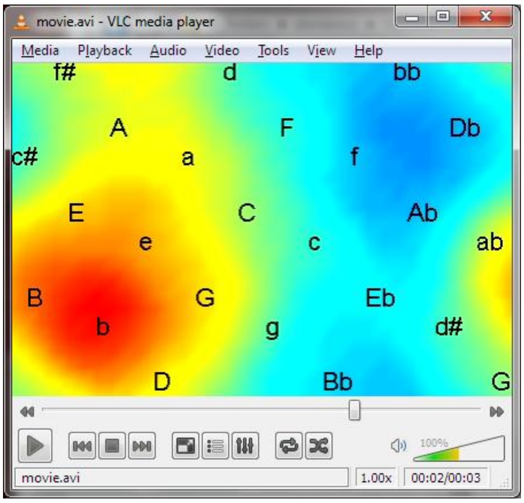

Figure 2 – Mapping EEG data to tonal maps (frame from animation)

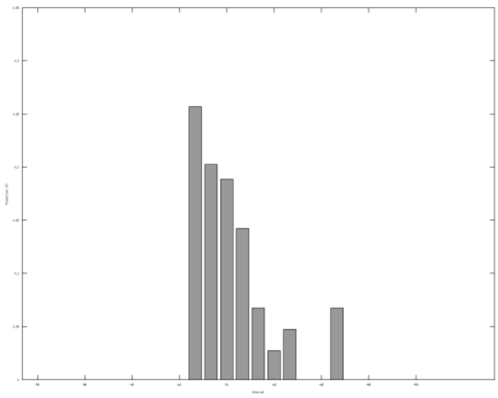

Figure 3 – Interval distribution analysis of EEG data by the automated analysis toolset

6. Conclusions

Sonification techniques are the most appropriate tools for extrapolating temporal-correlated properties in time series as musical patterns. In the context of epilepsy research, these musical patterns may hopefully be used as "markers" denoting specific epileptic EEG states, hence giving rise to the possibility to forecast an incoming seizure far in advance by just listening to how the musical patterns embedded in the on-going EEG change. The combination of recurrence quantification analysis and sonification technique is expected to deliver results and additional data better than canonical static representation of data (e.g. power spectra and related linear analyses), being intrinsically dynamic-oriented tools, hence suitable for depicting the intrinsic nonlinear nature of a dynamical disease as epilepsy

Biographies

Roberto Barbera graduated in Physics ‘cum laude’ from the University of Catania in 1986 and obtained a PhD in Physics from the same university in 1990. He is currently an Associate Professor at the Department of Physics and Astronomy of the University of Catania. Since 1997 he is involved with the ALICE Experiment at CERN/LHC and since 1999 he is interested in distributed computing. In 2004 he created the international GILDA Grid infrastructure for training and dissemination that he coordinates since then. He is the author of about 200 publications and more than 300 presentations at international workshops and conferences.

Dr. Francesca Falcetta was born in Gallarate (Italy) in May 1981. She graduated in Bioinformatics at the University of Milano Bicocca in 2005. She achieved P.h.D in Industrial Biotechnology in 2008. Since March 2009 she works in the department of Oncology at Istituto di ricerche farmacologiche Mario Negri in Milan. Her major research topic is the investigation of the effects of antitumoral drug on cell population.

Giuseppe La Rocca was born in Catania (Italy) in January 1975. He graduated in Software Engineering at the University of Catania in 2003. Since 2004 he joined the division of Catania of the Italian National Institute for Nuclear Physics (INFN) as a scientific collaborator and started to collaborate in several Grid-related projects co-funded by the

EC in the context of the 6th and 7th FP. Starting from 2006 he is the responsible for Grid deployment of the ASTRA project.

Massimo Rizzi, Ph.D., was born in Milan (Italy) in February 1963. He is actively involved in epilepsy research since 1989. He has recently turned his attention to the nonlinear analysis of EEG recordings of epilepsy patients in order to identify dynamical patterns potentially useful for seizure prediction and elucidate the pathophysiology of epilepsy according to the paradigm of the brain behaving as a complex system

Mariapaola Sorrentino graduated in electronic engineering (MScE) working in Cambridge, UK in the field of Quality Assurance, physical modelling and data sonification. She worked as test coordinator within the ASTRA (Ancient instrument Sound/Timbre Reconstruction Application) project, planning and coordinating the implementation of testing procedures to achieve the reconstruction of ancient musical instruments sounds, validating physical models and software algorithms. Mariapaola worked also as a lecturer at the Conservatory of Music of Avellino (Italy), bridging the gap between arts and technology, covering a wide range of topics from the fundamentals of computer science to the most advanced technologies of computer networking and multimedia.

Domenico Vicinanza works at DANTE, Cambridge, UK as a product manager. He received his MSc and PhD degrees in Physics and he is a professional music composer. He worked for several years as a Research Associate at University of Salerno and Roma Tre (Italy) and as a Scientific Associate at CERN (Geneva, Switzerland). His activities during this time included LHC Computing Grid sites management, IT resources management, network troubleshooting, grid computing and training/lecturing. He is also involved in the application of distributed computing and advanced networking technologies to music and visual arts as the technical coordinator of the ASTRA (Ancient instrument Sound/Timbre Reconstruction Application) and Lost Sounds Orchestra projects for the reconstruction of musical instruments on GÉANT and EUMEDCONNECT.