

BaSeBio: base de datos de señales biomédicas accesible desde Internet

C. Bonell¹, A.S. Cherniz², A. Hadad³, B. Drozdowicz⁴

^{1,2,3,4} Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina. Ruta Prov.11, Km. 10, Oro Verde, Dpto. Paraná, Entre Ríos, Argentina

e-mail: cbonell@bioingeniería.edu.ar

Resumen. En la actualidad, uno de los pilares que sustentan la aparición de productos médicos novedosos es el desarrollo de nuevas herramientas basadas en el procesamiento de señales obtenidas por diferentes métodos. Estas herramientas permiten extraer, a partir de estas señales, mayor información sobre la anatomía y los procesos fisiológicos subyacentes, a fin de arribar con mayor rapidez y en forma certera a un mejor diagnóstico médico. Por otro lado, el procesamiento de señales biomédicas ha posibilitado el conocimiento y la exploración de aspectos de la fisiología aún desconocidos. Para poder evaluar la seguridad y eficacia de los productos médicos que utilizan estas herramientas, documentar la investigación científica y posibilitar el intercambio y comparación de resultados entre grupos de investigación, es necesario contar con señales biomédicas registradas bajo condiciones conocidas, en sujetos seleccionados según criterios claramente establecidos y que los protocolos utilizados establecidos estén a disposición de quien haga uso de estas señales. Por esto, el presente trabajo presenta el diseño y desarrollo de un sistema para la gestión de un banco de datos de señales biomédicas, basado en Internet por medio del cual será posible acceder en forma libre a señales e imágenes biomédicas.

1. Introducción

En la actualidad, el desarrollo de nuevas tecnologías en la industria médica y la investigación aplicada en el área de la bioingeniería se sustenta, en gran parte, en el desarrollo de nuevos algoritmos de procesamiento de señales. La evaluación del desempeño de estos algoritmos debe incluir la validación de los resultados obtenidos a partir del análisis y procesamiento de señales reales (señales electrofisiológicas o imágenes médicas) y la evaluación de su desempeño ante diferentes condiciones de registro. Esta validación es un requisito fundamental establecido por las regulaciones y normas nacionales e internacionales para productos médicos [1,2]. Esta etapa es la más difícil, fundamentalmente debido a que las señales biomédicas suelen tener una gran variabilidad entre sujetos e, inclusive, en el mismo sujeto en distintas instancias de registro [3,4]. Además, las características propias de las señales dependen de las condiciones en las cuales han sido registradas y requieren muchas veces de la utilización de estrategias de normalización [5].

La validación de los resultados obtenidos debe incluir la utilización de herramientas estadísticas adecuadas para evaluar la consistencia de los mismos [6,7]. Esto trae aparejado otro inconveniente, además de los ya mencionados, que es disponer de cantidad suficiente de conjuntos de datos estadísticamente representativos para los posibles escenarios de adquisición. Por esto, muchas de las aplicaciones clínicas de la mayoría de los algoritmos y procesamientos desarrollados en el ámbito de

la investigación son limitadas [8]. En las primeras etapas del desarrollo de nuevos algoritmos, suelen utilizarse modelos, pero el planteo de los mismos está condicionado por el conocimiento de los mecanismos fisiológicos (muchas veces limitado), la variedad de fenómenos físicoquímicos, la interdependencia entre los distintos sistemas biológicos y la presencia de procesos no lineales y situaciones no estacionarias, presentes en la génesis de las señales biológicas [9-12]. A esto hay que agregar que, en determinadas situaciones, no existe un patrón o método de contraste que permita medir directamente el desempeño del algoritmo, lo cual hace aún más difícil la evaluación de los resultados obtenidos. Es por esto que la validación del software por medio de modelos es solo el primer paso y debe ser continuado con la evaluación en señales reales obtenidas en distintos sujetos.

A fin de poder comparar los resultados de distintas herramientas de procesamiento y evaluar los dispositivos que utilicen estas herramientas para el diagnóstico o como parte de una nueva tecnología de la industria médica, es necesario contar con un número suficiente de señales e imágenes tomadas en condiciones de registro controladas, confiables y reproducibles. Por esta razón, en este trabajo se presenta el desarrollo de una base de datos de señales e imágenes médicas. La misma ha sido pensada como respaldo para las tareas de prueba y validación de nuevos algoritmos, de forma tal que permita la documentación de resultados científicos, comparación de resultados y comunicación entre distintos grupos de investigación que utilicen estas señales.

1. Materiales y métodos

Para el desarrollo de la base de datos se utilizaron herramientas de software libre. El motor de la base de datos utilizada es el MySQL, el cual posee licencia de uso gratuita y permite el manejo de volúmenes de información moderadamente grandes. Además cuenta con medidas de seguridad que permiten la administración de privilegios para el acceso a la información por parte de distintos tipos de usuario (administrador, laboratorio, invitado – ej.: investigador ó estudiante). Por otro lado es posible alojar la base de datos en cualquier plataforma (Windows, Linux, etc.) y contiene mecanismos que facilitan el respaldo y restauración de la información alojada mediante copias de seguridad.

1.1. Diseño del sistema de gestión de la base de datos

El sistema de gestión de la base de datos comprende tres módulos: uno destinado a la carga de nuevas señales, otro correspondiente a la autorización de las cargas realizadas y verificación de los datos de las mismas y un tercero para la consulta y acceso a las señales contenidas en la base.

El *módulo de carga* permite el ingreso de nuevas señales o imágenes a la base. Para ello, el usuario debe estar registrado como usuario del tipo *laboratorio*. Esto quiere decir que, cada laboratorio autorizado para añadir registros contará con un usuario y contraseña especiales, suministrado por el usuario del tipo *administrador*. Todos los integrantes de un laboratorio necesitarán contar con el usuario y contraseña provistos para acceder al *módulo de carga*. Para que la carga de una señal sea válida, además de estas, se deben ingresar al sistema los siguientes datos: código del paciente, fecha del registro, laboratorio (del listado de laboratorios registrados), tipo de señales que contiene el registro (mediante los acrónimos autorizados), número de protocolo y número de sesión. Esto se realiza a través de la pantalla de carga, la cual contiene campos específicos para tal fin. Los datos ingresados se almacenan en la base de datos y se utilizan para la composición de un identificador único ó código alfanumérico para cada registro.

Además de los datos mencionados, la base de datos guarda información de la dirección física de un archivo comprimido que contiene el protocolo de registro y las señales propiamente dichas. El archivo se genera en forma automática durante la carga de datos luego de la selección de archivos de señal y protocolo elegidos por parte del usuario en su estación de trabajo local. El nombre del archivo comprimido será de la forma del código alfanumérico antes descripto. El primero de estos archivos corresponde al protocolo, debiendo ser del tipo texto con extensión .txt o .doc. En él se establece la cantidad de señales registradas y el tipo de señal, como así también las condiciones de registro: tipo de paciente, rango de los filtros, forma de adquisición, etc. Cada laboratorio autorizado para la carga deberá guardar registro de los protocolos que le son propios y deberá ingresar su identificación al

momento de la carga de señales. Los otros archivos corresponden a las señales propiamente dichas, y podrán contemplar uno o más archivos, de acuerdo con el formato establecido y descripto en el protocolo.

El módulo destinado a la verificación de datos sólo puede ser visualizado por los usuarios que tienen la categoría de *administrador*. La función del administrador es verificar que al momento de la carga se hayan consignado todos los datos que se solicitan en la pantalla de carga, que los archivos cargados contengan el protocolo y que los datos contenidos en el mismo sean los necesarios para la correcta interpretación de las señales y sean consistentes a los datos ingresados en el módulo de carga. Para ello el sistema informa al administrador la existencia de nuevos datos. Una vez revisados, y si corresponde, procede a autorizar la carga en la base de datos propiamente dicha. En caso de inconsistencias o errores, se rechaza el registro y se devuelve al laboratorio que subió los datos un informe donde constan los motivos de tal decisión.

Las señales incluidas en la base podrán ser consultadas a través del *módulo de consulta*. De tal manera, todos los usuarios registrados pueden descargar el archivo comprimido conteniendo el protocolo y las señal/es que sean de su interés.

En relación a estos módulos cada tipo de usuario posee distintos clases de permiso. Para componer dichos permisos se generan roles que permiten acceder a las acciones propias de cada uno de ellos. Así, entonces, el usuario registrado como *laboratorio* tiene el rol de *carga*, con lo cual tiene acceso al *módulo de carga* previamente descripto. También podrá acceder a la información relativa al historial de cargas y los informes de rechazo, si los hubiere. El rol de *autorizador* se reserva al *administrador* de la base quien aprueba o rechaza la incorporación de datos a la misma. Este *administrador* puede además, mediante funciones adicionales, generar acrónimos para la identificación de las señales en la base y aprobar nuevos usuarios y roles de los mismos. El rol de *consulta* se reserva a todos aquellos usuarios que deseen obtener señales de la base. En la figura 1 se representa el diagrama de actividades antes descripto y los roles y actividades de cada tipo de usuario.

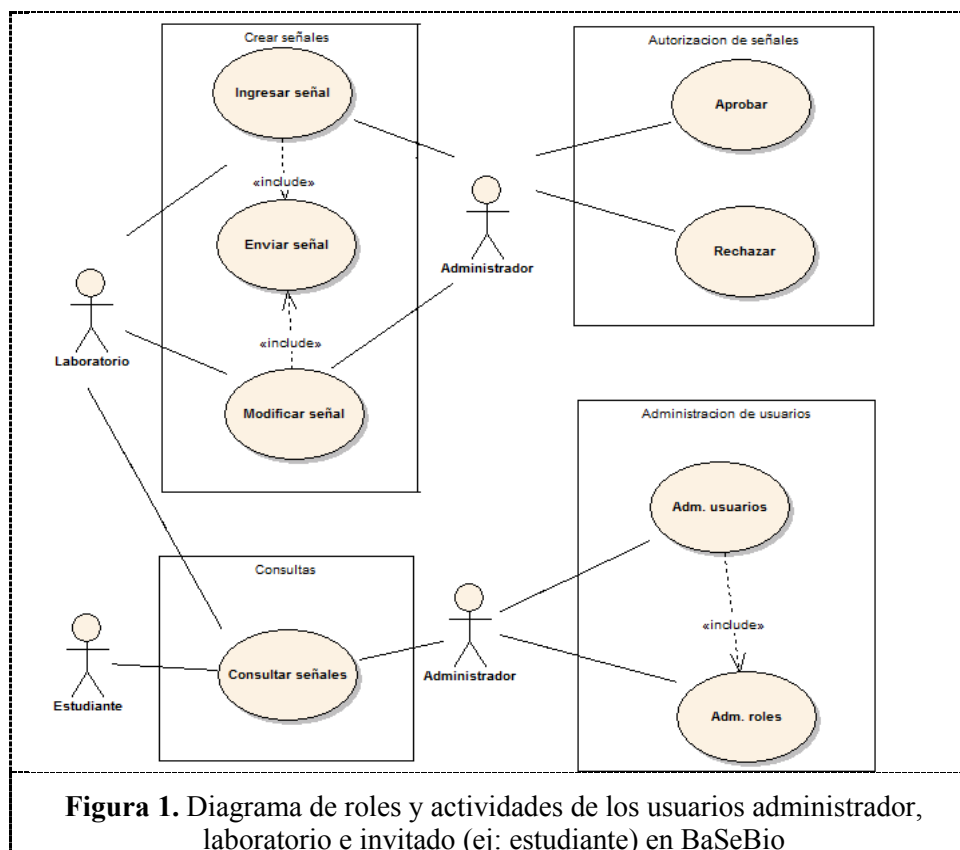


Figura 1. Diagrama de roles y actividades de los usuarios administrador, laboratorio e invitado (ej: estudiante) en BaSeBio

1.2. Información contenida en la base

Cada archivo comprimido, generado durante la carga de señales en la base de datos (por parte de los usuarios con la categoría de *laboratorio*), deberá contener el protocolo utilizado para la generación de los datos y los archivos de datos correspondientes. Para ello, la base de datos contiene un archivo de ayuda en el cual se detallan los aspectos que deberán ser consignados en el protocolo. Estos son: título del protocolo, materiales y métodos, registro de las señales (incluyendo los datos necesarios para su recuperación y su formato) y opcionalmente el modelo de consentimiento informado. Podrá incluirse, si correspondiere, un archivo de marcadores que describan indicadores en la señal a intervalos de tiempos particulares.

Adicionalmente a esta información, la base contiene un listado de señales con sus acrónimos y descripción de forma tal de unificar su identificación. Es posible obtener, además, el listado de los protocolos generados (código alfanumérico asignado para el laboratorio que lo diseña) y una breve descripción del mismo. En la figura 2 se muestra el Diagrama Relacional en el cual se observa cómo se modelaron los requerimientos derivados del diagrama de Roles y Actividades (figura 1). Las entidades fundamentales que se representaron son las asociadas a los laboratorios y a los responsables de los mismos que intervienen en la carga de señales e imágenes y los datos asociados a las mismas.

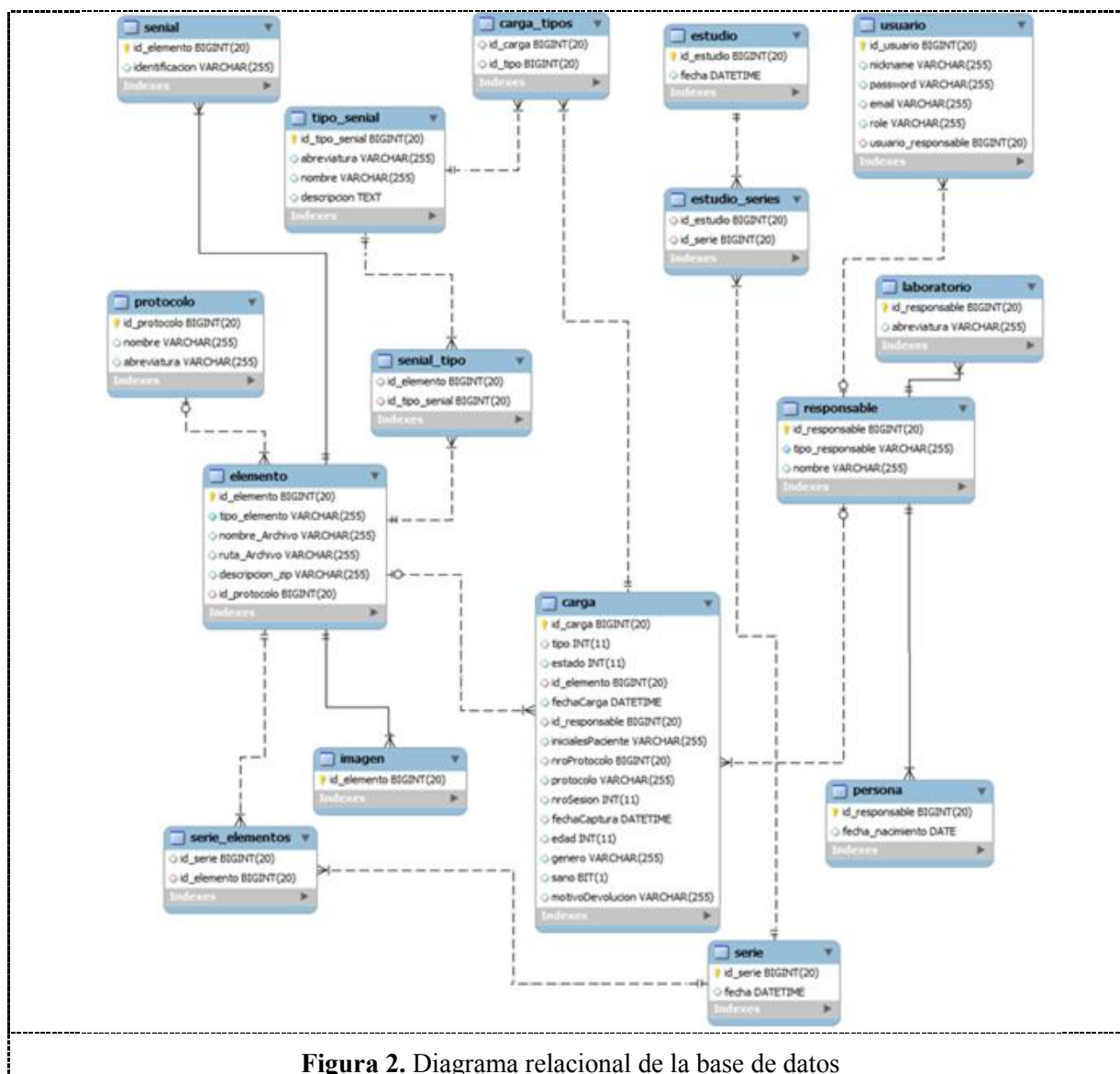


Figura 2. Diagrama relacional de la base de datos

1.3. Implementación de la base

Como etapa final del proyecto, se procedió al desarrollo de la aplicación para el ingreso de información referida a señales e imágenes médicas. Dicha aplicación fue desarrollada bajo el entorno web del lenguaje JAVA, entre cuyas ventajas se cuenta la posibilidad de implementar los desarrollos bajo diferentes entornos (Sistemas Operativos) y que posee diversos complementos (frameworks) que facilitan la tarea de desarrollar el software, como por ejemplo, los disponibles para el manejo de la persistencia de los datos, la presentación de la información o la administración de la seguridad. La aplicación consta de los módulos descritos en la sección 1.1. Los complementos de JAVA utilizados fueron Hibernate para la administración de la base, que posibilita el manejo y la persistencia de los datos y Wicket para la presentación de la información, a fin de facilitar el manejo de los diferentes componentes de las pantallas donde el usuario interactúa con el sistema. Por último se utilizó Acegi para administrar los diferentes permisos de los usuarios y asignar los roles que les brindan las autorizaciones necesarias para realizar cada una de las actividades dentro del sistema. El sistema de gestión de la base de datos comprende tres módulos: uno destinado a la carga de nuevas señales, otro correspondiente a la autorización de las cargas realizadas y verificación de los datos de las mismas y un tercero para la consulta y acceso a las señales contenidas en la base.

2. Resultados

La base de datos se instaló en el servidor de la Facultad de Ingeniería de Universidad Nacional de Entre Ríos. El programa posee en su inicio una interfaz que permite acceder a la base vía Internet. De acuerdo al nivel de acceso que se posea, previo ingreso del nombre de usuario y clave, se accede a la configuración del menú correspondiente.

En la figura 3 se muestra el aspecto de la página de inicio de **BaSeBio**, bajo el rol de administrador. Bajo este rol es posible cargar señales y, tal como se muestra en las pestañas del menú superior, autorizar y/o devolver señales generadas por otros laboratorios. El menú *administración* permite generar nuevos usuarios con la capacidad de generar datos para la base o solo para su consulta.



En la figura 4 se muestra el aspecto del menú de carga de señales. Por medio de este menú se agregan señales a la base, pudiéndose seleccionar entre imágenes y señales electrofisiológicas. Para tal fin, es necesario completar los datos correspondientes al módulo de carga. Una vez ingresados estos datos se solicita seleccionar los archivos que contienen el protocolo de registro, las señales y los marcadores (si corresponde). Estos archivos conformarán el archivo comprimido (.zip) cuyo enlace se almacenará en la base de datos para su recuperación.

Figura 4. Aspecto de la ventana para el menú de carga de señales, con el rol de “administrador” o “laboratorio”.

La figura 5 muestra el aspecto del menú de consultas. La búsqueda de información se puede realizar utilizando el tipo de señal o señales de interés (seleccionable de una lista de disponibles), texto contenido en el nombre del protocolo, código de paciente, fecha de captura, sesión o laboratorio que generó la señal. Una vez seleccionados los criterios de búsqueda y realizada la misma, se muestran en pantalla el listado de todas aquellas señales contenidas en la base que cumplen con los criterios establecidos. Del listado se seleccionan aquellos que sean de interés y se bajan los archivos comprimidos que contienen el protocolo y las señales generadas.

Tipo	Paciente	Protocolo	Sesión	Fecha	Responsable	Estado	
[EMG]	pepe arias	dfds	3	03/03/2011	Administrador	Pendiente	[ZIP]
[EEG]	samuel	jkuioo	2	12/04/2011	Administrador	Pendiente	[ZIP]

Figura 5. Aspecto de la ventana para el menú de consulta.

Luego del desarrollo, se llevaron a cabo pruebas para evaluar el correcto funcionamiento, llegándose a resultados aceptables para el desempeño del mismo, tanto en tiempos de respuesta, como resguardo de la información o resolución de errores en el funcionamiento.

3. Discusión y conclusiones

El sistema de gestión de la base de datos de señales presentado en este trabajo surgió a partir de la necesidad de contar con señales electrofisiológicas e imágenes médicas registradas en condiciones específicas para el desarrollo y validación de algoritmos. Los protocolos de adquisición de las señales disponibles en los bancos de datos públicos en numerosas ocasiones no presentan las características deseadas, ya sea por las condiciones en que se registran las señales, la ausencia de marcadores de eventos, la cantidad necesaria de señales para la evaluación estadística o las características de los sujetos. Bases de datos como PhysioBank [13-14] son ampliamente utilizadas por la comunidad científica para la evaluación del desempeño de nuevos algoritmos. Esta base de datos contiene varios tipos de señales, siendo su mayoría señales de electrocardiografía y de electroencefalografía. Otras bases de datos libres se restringen solo a señales de electrocardiografía (ECG) [15] o están destinadas a discutir la resolución de determinados problemas específicos, como puede ser la evaluación del desempeño de algoritmos para la descomposición de señales de EMG [16].

A partir de esta necesidad se desarrolló el sistema de gestión BaSeBio el cual no sólo da respuesta a esta demanda de datos, sino optimiza la generación, protocolización y almacenamiento de señales e imágenes de forma tal de garantizar la calidad y cantidad de los datos almacenados. El sistema de gestión desarrollado ha resultado una herramienta versátil para el almacenamiento tanto de señales electrofisiológicas como de imágenes médicas, centralizando en una sola aplicación el almacenamiento tanto de señales unidimensionales como bidimensionales. Por otro lado, la necesidad de documentar los protocolos de registro de las señales y mantener el registro de estas actividades para su ingreso a la base de datos, optimiza la utilización de estas señales en múltiples instancias, tanto de investigación como de docencia. Tanto las señales registradas como los protocolos una vez incorporados a la base por el usuario que los genera, son revisados por el administrador de la base lo cual busca garantizar la calidad y recuperabilidad de la información. Adicionalmente, el sistema provee de un sistema de búsqueda que puede realizarse por varios criterios, con el fin de facilitar el acceso a la información tanto por parte de usuarios externos como internos de la FIUNER. Con este sistema se espera contribuir a la validación y contrastación de nuevos desarrollos poniendo las señales a disposición de la comunidad. Además, se provee a futuro incorporar software de análisis de señales, herramientas de simulación y resultados obtenidos en las investigaciones que las utilicen.

Agradecimientos

La programación de la base de datos fue financiada por el proyecto “Tecnologías para el desarrollo, validación y determinación de seguridad eléctrica de productos médicos de Entre Ríos” del “Primer Programa de Fortalecimiento de la I+D y la Ciencia en la Sociedad” de la Agencia de Ciencia, Tecnología e Innovación de Entre Ríos (ACTIER).

Referencias

- [1] Donawa, M. 2002 *Med.Device Technol.* **13** 20.
- [2] Keatley, K. L. 1999 **7** 49.
- [3] Dorel, S., Couturier, A., and Hug, F. 2008 **18** 857.
- [4] Hug, Francois, Turpin, Nicolas A., Gu+@vel, Arnaud, and Dorel, Sylvain 1-6-2010 *Journal of Applied Physiology* **108** 1727.
- [5] Rouffet, D. M. and Hautier, C. A. 2008 **18** 866.
- [6] Carter, R. E. 13-1-2005 *BMC.Med.Res.Methodol.* **5** 3.

- [7] Sayee Sundar, K., Meiyappan, S., and Prakash, O. 1-4-1989 *International Journal of Clinical Monitoring and Computing* **6** 127.
- [8] Meekins, Gregg D., So, Yuen, and Quan, Dianna 2008 *Muscle & Nerve* **38** 1219.
- [9] Beauchet, O., Dubost, V., Herrmann, F. R., and Kressig, R. W. 2005 *J.Neuroeng.Rehabil* **2** 26.
- [10] Goldberger, A. L. and West, B. J. 1987 *Ann.N.Y.Acad.Sci.* **504** 195.
- [11] Goldberger, A. L., Amaral, L. A., Hausdorff, J. M., Ivanov, P. C., Peng, C. K., and Stanley, H. E. 19-2-2002 *Proc.Natl.Acad.Sci.U.S.A* **99 Suppl 1** 2466.
- [12] Scafetta, Nicola, Marchi, Damiano, and West, Bruce J. 2009 **19** 026108.
- [13] Costa, M., Moody, G. B., Henry, I., and Goldberger, A. L. 2003 *J.Electrocardiol.* **36 Suppl** 139.
- [14] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., and Stanley, H. E. 13-6-2000 *Circulation* **101** E215-E220.
- [15] Couderc, J. P. 2010 *J.Electrocardiol.* **43** 595.
- [16] Florestal, J. R., Mathieu, P. A., and McGill, K. C. 2009 **19** 1.