

Vorschlag für ein Dateiformat für die Verarbeitung, die Archivierung und den Austausch von Biosignal-Daten

MARKUS KUHN⁽¹⁾, MARKUS PROSCH⁽¹⁾

(1) Institut für Physiologie und Biokybernetik, Universität Erlangen-Nürnberg, Universitätsstr. 17, D-91054 Erlangen

Digitalisierte Meßdaten aus Aufzeichnungsgeräten für Biosignale (EEG, MEG, ECoG, EKG, usw.) werden derzeit an vielen Institutionen in einer oft unüberschaubaren Zahl von verschiedenen Formaten verarbeitet und aufbewahrt. Viele für Biosignale interessante Algorithmen und Programme wie etwa zur Filterung, Visualisierung, Merkmalsextraktion, Korrelation, statistischen Analyse, usw. lassen sich weitgehend unabhängig vom benutzten Meßgerät und der Signalherkunft realisieren. Dennoch werden derartige Programme derzeit eher selten für Daten aus anderen Quellen als denen verwendet, für die die Software ursprünglich vorgesehen war. Es werden auch oft nur rohe Meßwerte ohne jegliche Begleitinformation abgespeichert. Dies erschwert die Archivierung, da eine Beschreibung der Daten getrennt abgelegt werden muß. Zusätzlich ist dies eine mögliche Fehlerursache, da die Software selbst keinen Zugriff auf für die Analyse wichtige Parameter (Abtastrate, Skalierungsfaktoren, physikalische Einheit, usw.) hat und diese vom Benutzer manuell angegeben werden müssen. Es sind zahlreiche herstellerspezifische Formate in Gebrauch, die auf sehr spezielle Anwendungen zugeschnitten und oft nur unzulänglich dokumentiert sind, was die Verarbeitung, den Austausch und die Archivierung der Daten erschwert.

Daher wurde das Dateiformat EBS (*extensible bio-signal file format*) entworfen. Nach dem Vorbild der Entwicklung in anderen Bereichen der angewandten Informatik wie etwa in der Computergraphik, wo einheitliche Dateiformate schon vor Jahren zur leichten Austauschbarkeit von Daten und Werkzeugen geführt haben, ist zu hoffen, daß dies auch im Bereich der Biosignal-Analyse möglich ist.

Das EBS-Format ermöglicht die Aufnahme von mehreren synchron aufgezeichneten Kanälen, auch aus verschiedenen Signalquellen, in eine Datei (gleiche Abtastrate vorausgesetzt). In beliebig vielen sogenannten Attributen können weitere Informationen über die Daten in der selben EBS-Datei mit abgelegt werden. Es wurde bereits eine Reihe von Attributen für verschiedene Anwendungsgebiete definiert, um eine einheitliche Basis herzustellen, jedoch existieren auch Mechanismen, die die einfache, kollisionsfreie und individuelle Erweiterung für beliebige neue und spezielle Attribute erlauben.

Zu den vordefinierten Attributen gehören Informationen über:

- den Patienten/Probanden (Name, Geburtsdatum, Kennung, Geschlecht),
- die Aufzeichnung (Benennung und Beschreibung der Kanäle, physikalische Maßeinheiten der Kanäle, eingesetzte Filter, Abtastrate, graphische Darstellung der Sensor- bzw. Elektrodenpositionen, Zusammenfassung der Kanäle zu Gruppen, Markierung von Ereignissen, Aufnahmezeitpunkt),
- die Institution, an der die Aufzeichnung oder Verarbeitung durchgeführt wurde,
- sowie eine Auflistung der bisherigen Verarbeitungsschritte,
- eine einzeilige Kurzbeschreibung des Dateiinhaltes für Übersichtslisten und
- ein beliebig langer ausführlicher Beschreibungstext.

EBS-Dateien können Meßdaten in verschiedenen Formaten enthalten. Derzeit werden ganzzahlige vorzeichenbehaftete 16-bit Werte abgelegt, aber die zusätzliche Unterstützung anderer Wortgrößen oder Fließkommawerte ist vorgesehen. Während Aufzeichnungsgeräte in der Regel die Daten nach dem Samplezeitpunkt sortiert ablegen, da sie in dieser Reihenfolge anfallen, ist es für Analyseprogramme, die sich mit einzelnen Kanälen befassen, oft günstiger, die Daten nach Kanälen gruppiert zu erhalten. Im EBS-Format werden beide Alternativen unterstützt und können durch Konvertierprogramme leicht ineinander überführt werden. Es besteht ferner die Möglichkeit, durch Abspeichern der Differenz zum jeweils vorangegangenen Meßwert im selben Kanal für die Mehrzahl der Meßwerte mit weniger Datenbits auszukommen. Dadurch läßt sich die Dateilänge ohne Informationsverlust typischerweise fast halbieren. Diese *lossless compression* eignet sich insbesondere zur kostengünstigen Archivierung. Weitere leistungsfähigere Kompressionsoptionen für EBS werden derzeit untersucht.

Sehr universelle Formate werden wegen des hohen Implementationsaufwandes im wissenschaftlichen Bereich nur selten eingesetzt. Beim Entwurf des EBS-Formates wurde daher darauf geachtet, daß es leicht verständlich bleibt und Zugriffsprozeduren in kurzer Zeit erstellt werden können. Um dieses Entwurfsziel zu erreichen, wurde auf komplexere Möglichkeiten wie etwa die Unterstützung von unterschiedlichen Abtastraten oder Datenformaten in der selben Datei verzichtet. Internationale Normen wurden berücksichtigt, sofern dies die Komplexität nicht unnötig erhöhte. Insbesondere wird bei den Attributen ein moderner 16-bit Zeichensatz (ISO 10646) verwendet, so daß nicht nur Englisch, sondern beliebige Sprachen wie etwa Japanisch oder Arabisch in Beschreibungstexten verwendet werden können. Bei der Entwicklung des EBS-Formates wurde den Möglichkeiten moderner Betriebssysteme und Prozessoren Rechnung getragen (*32-bit alignment*, *memory mapping* möglich, usw.), so daß eine effiziente und komfortable Verarbeitung auf modernen Workstations ebenso möglich ist wie der Einsatz auf preiswerten Personal Computern.

Eine EBS-Datei besteht aus 3 oder 4 Abschnitten: Sie beginnt mit einem festen 32 Byte langen Datenblock (*fixed header*) mit Informationen wie Anzahl, Länge und Kodierung der Kanäle, die von allen Programmen benötigt werden. Im anschließenden Teil (*variable header*) sind die Attribute mit jeweils einer Kennung und einer Längenangabe abgelegt, so daß unbekannte Attribute leicht übersprungen werden können. Attribute können auch nach dem eigentlichen Datenteil in einem zweiten *variable header* folgen. Dies ermöglicht den einfachen Umgang mit sehr großen Dateien, da Attribute, die nachträglich hinzukommen oder verlängert werden (z.B. nachträgliches Hinzufügen einer Beschreibung) und daher nicht mehr vor den Datenteil passen, an das Ende der Datei angehängt werden können. Dadurch wird bei Änderungen ein sehr zeit- und speicherplatzaufwendiges Umspeichern vermieden. Damit der Zugriff auch auf Attribute wachsender EBS-Dateien (z.B. während der Aufzeichnung) möglich ist, können diese auch vor den Meßdaten im ersten *variable header* gespeichert werden.

EBS-Software im ANSI C Quellcode für UNIX- und MS-DOS-Rechner ist bereits verfügbar bzw. in Vorbereitung. Diese, sowie die komplette englischsprachige Spezifikation des Formates ist bei den Autoren erhältlich oder über das Internet mit *anonymous ftp* vom Server [ftp.uni-erlangen.de](ftp://ftp.uni-erlangen.de) im Verzeichnis `pub/eps/` zu finden.

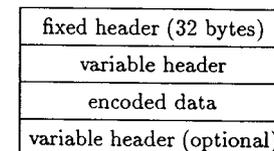


Abb. 1: Aufbau einer EBS-Datei