



Working Together: An EMG (Decomposition) User's Group

XVIIth Congress of the International Society for
Electrophysiology and Kinesiology

Workshop: EMG Decomposition



Copyright Edward A. Clancy and Kevin C. McGill, 2008. Some rights reserved.
Content in this presentation is licensed under a Creative Commons Attribution-
Noncommercial-Share Alike 3.0 License. This license is more fully described at
<http://creativecommons.org/licenses/by-nc-sa/3.0/>

EAC08-100

Purpose of a User's Group



- Develop **SHARING** of data, software, etc.
 - Data, annotation formats
 - “Blue Ribbon” annotations, comparison
 - Decomposition accuracy (overall & individual MUAPT)
 - Post-processing of decomp data
 - Quality control for all the above!
 - IRB, intellectual property issues

EAC08-101

Why Share Data, Software, Etc.? (“Pros”)



- Avoid repeated efforts
- Gain from research of others
- Spend time on research advances, not incorporating past knowledge
- Advance collaboration (virtual labs)
- Develop community that shares well beyond data, software
- Get others to use, cite your research

EAC08-102

Data/Software Sharing Challenges (“Cons”)



- Do I trust the quality of the work?
- Will other data/software fit my needs?
- Lack of standards to exchange data, annotations (Some exist for data!)
- Different processing environments
 - MATLAB is common to many
- Much harder to prepare a shared tool
 - Documentation, tool organization, ease of use by others
- Credit for my work?
- Will others use my work competently?
- Proprietary concerns (for businesses & academics!)

EAC08-103

Who Should Participate?

- Contributors
 - Any who agree to open source model
 - Academic & commercial
- Users
 - Open to all, including those NOT contributing!
- Q: What makes contribution beneficial?






EAC08-104

Biomed Sharing Examples

- PhysioNet (<http://www.physionet.org>)
 - Shared database, software, annotations, etc.
 - Primarily for ECG (but, others also)
 - Generated hundreds of research projects
- EEGLAB (<http://www.sccn.ucsd.edu/eeelab/>)
 - Interactive MATLAB toolbox for EEG, MEG
 - User extensible
 - 2500+ researchers in e-mail discussion list
- Neuroshare.org (<http://neuroshare.org>)
 - ENG spike sorting. Open-source, commercial MATLAB, C.
- Neurodatabase (<http://neurodatabase.org>)
 - Public basic neuroscience database, db tools
- Brain Resource International Database
 - http://www.brainnet.org.au/research/pub_guide.html
 - <http://www.brainresource.com/>
 - EEG data from 50+ labs
- Australian EEG database (<http://esg.newcastle.edu.au/inquiry>)
 - 18,500 EEG records. Click for [descriptive Journal paper](#).
- Numerous commercial standards

Open Access Journals
e.g. BioMed Central's "Journal of
Neuroengineering and
Rehabilitation"
(<http://www.neuroengrehab.com/home/>)

EAC08-105

Biomed Sharing Resources

- U.S. NIH data sharing policy:
 - http://grants.nih.gov/grants/policy/data_sharing/
- Cornell Lab of Neuroinfo (Daniel Gardner)
 - <http://neurodatabase.org/>, <http://datasharing.net/>

Sharing Data

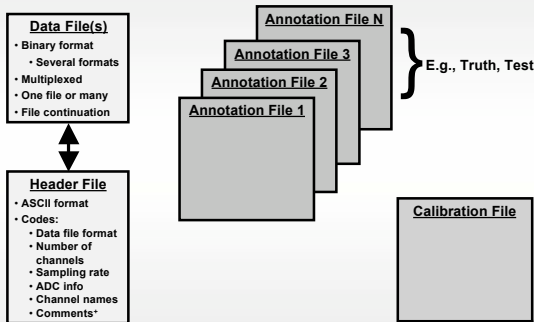
Data Storage Format

- Several formats already exist
- Chose to separate data from annotations
- Selected WaveForm DataBase (WFDB) format for EMGLab
 - Binary data files (many formats)
 - Associated header, annotation files

WFDB Origins

- George Moody, Roger Mark at MIT/BIH
- Used extensively in ECG, clinical monitoring
- Developed in 1980s (??)
 - Lot of C-software support
 - Original database: ECG Holter
- PhysioNet standard (<http://physionet.org>)
 - Originally NIH funded
 - Open source; numerous databases

WFDB Structure (For Each Trial)



WFDB Binary Formats

- 8: 8-bit first difference
- 16: 16-bit, 2's complement, little endian
- 61: 16-bit, 2's complement, big endian
- 80: 8-bit, offset binary
- 160: 16-bit, offset binary, little endian
- 212: 12-bit, 2's complement, packed
- 310: 10-bit, 2's complement, packed
- 311: 10-bit, 2's complement, packed

Header File Example

```

Record Line: Name, chans, F_samp
R00901 4 25000
R00901.dat 16 3276800(0)/Volts 16 0 0 0 Channel 1
R00901.dat 16 3276800(0)/Volts 16 0 0 0 Channel 2
R00901.dat 16 3276800(0)/Volts 16 0 0 0 Channel 3
R00901.dat 16 3276800(0)/Volts 16 0 0 0 Channel 4
# Quadrifilar needle recordings.
# First dorsal interosseous muscle, healthy subject.
# Channel_1 = Contact_A - Contact_B.
# Channel_2 = Contact_B - Contact_C.
# Channel_3 = Contact_C - Contact_D.
# Channel_4 = Contact_A - Contact_D.
# Hardware: 300 Hz high pass (passive, first order),
# 10 k Hz low pass (4th-order, Butterworth),
# Gain = 500.
# +5 V ADC, 16 bit.
# Trapezius force profile:
# zero to 20% MVC to zero.

```

Labels in diagram: Record Line, Name, chans, F_samp, Filename, Comment Lines, Gain(Offset), Format, ADC Resolution, Description, Signal Lines.

VA PAHCS EAC08-112 WPI

IRB Issues for Data Contributors

- IRB: Institutional Review Board
- **ALWAYS** subject to local IRB practices, permissions
- **Best** → describe data archiving in written Informed Consent document
 - Else, IRB re-approval, subject re-consent may be needed
- **De-Identify** all data before contribute
 - Remove name, initials, address, SS#, etc.

VA PAHCS EAC08-113 WPI

U.S. NIH: “Coded” vs. “Unlinked”

- **Coded data**
 - Subject identities replaced with codes
 - Original investigator holds “key”
 - Key relates codes to identities
 - U.S.: Requires IRB supervision (Exemption)
- **Unlinked data**
 - Subject identities replaced with codes
 - Key non-existent or destroyed
 - Can *never* determine identities
 - U.S.: Not “human subject.” **IRB Not Applicable**

ALWAYS confirm with local IRB !!!

VA PAHCS EAC08-114 WPI

U.S. IRB Resources

- “Some Human Studies Considerations for Potential Data Contributions to the EMGlab Website” (<http://emglab.stanford.edu/EMGLAB/Contribute/IRB.html>)
 - Includes suggested wording for IRB application, Informed Consent document
- “NIH Requirements for the Research Use of Stored Human Specimens and Data” (<http://ohsr.od.nih.gov/info/sheet14.html>)
- “Guidance on Research Involving Coded Private Information or Biological Specimens” (<http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf>)
- “Research Use of Stored Human Samples, Specimens or Data” (http://ohsr.od.nih.gov/info/DDIR_memo.html)
- “Points to Consider in Development of Informed Consent Documents that Include the Collection and Research Use of Human Biological Materials” (<http://ohsr.od.nih.gov/info/sheet15.html>)

VA PAHCS EAC08-115 WPI

Sharing Annotations

VA PAHCS EAC08-116 WPI

Annotations: Introduction

- **Annotations** = spike firing times, classification results, etc.
 - A.k.a.: spike files, detection/classification results
- **Recommend distinct file from data**
 - Data are fixed, unchanging
 - Annotations can evolve, change
 - Can have multiple annotation files per data file, e.g.:
 - Automated
 - Manually corrected
 - Annotate all spikes or dominant spikes
 - Regular spikes or clinical abnormalities
 - Spike times or other information (e.g., artifact, experimental intervention)

VA PAHCS EAC08-117 WPI

Existing Annotation Formats

- **Surveyed (2004+):**
 - Englehart (University of New Brunswick)
 - Erim (Rehab Institute of Chicago)
 - Farina (SMI, Aalborg University)
 - Kamen (UMass-Amherst)
 - McGill (Palo Alto VA / Stanford University)

- **Common elements:**
 - 1) Firing time
 - 2) Spike ID

- **Additional elements:**
 - Too many to list!!



Existing Annotation Formats: Summary

- **Many existing formats**
- **Varying information preserved**
 - Each format stores spike times, identities
- **Existing *formats* conflicting**

No existing format could be used to satisfy every user

Annotation Format: What do we need?

- Usable in MATLAB, but NOT proprietary
- Preserves legacy information
- Requires *firing time* and *spike ID*
- Permits many other parameters
- Extendable (new users, future uses)
- Facilitates standardized performance comparison, post-processing



Flexible format required

Binary Format vs. ASCII

- **Usually: “ASCII is for amateurs!”**
 - E.g.: Never store ADC data in ASCII
- **But:**
 - Annotation info much smaller size than data
 - ASCII very flexible, extensible, not machine-specific, variable length lines
 - Existing information standards, e.g. XML

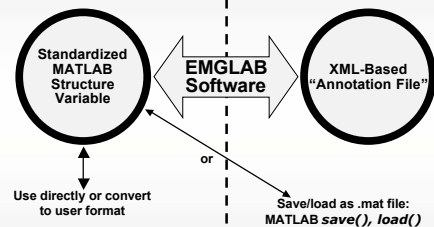
Pros/Cons of Annotation Standards

- **Cons**
 - Inefficient if only 1–2 users
 - Satisfy all your needs ????????
 - Software organization/style concerns
- **Pros**
 - Very efficient if exchange with many researchers ($2 \cdot N$ instead of N^2)
 - Facilitates sharing !!!
 - Accompanying documentation
 - Supportive of databases

Annotation Format: Overall Solution

MATLAB Environment

Share With Others



EMGLAB Annotation *Structure*

- MATLAB-based
- Three information categories
 - General info: General to entire decomp
 - Spike events: One “row” of info per spike
 - Other: E.g., MU templates, spike epochs
- Three required fields
 - Version number
 - Spike time, ID (per event)



EMGLAB Annotation *File*

- XML (eXtensible Markup Language)
- Eight sections
 1. XML declaration statement
 2. Root element
 3. Version tag
 4. Optional general information
 5. Spike header tags
 6. Spike events
 7. Optional additional structure fields
 8. Optional freeform variables

} One-to-one correspondence with EMGLAB annotation structure

EAF: Simple Example

```

1) Declaration statement <<?xml version="1.0" encoding="ASCII"?>
<emglab_annotation_file>
  3) Version tag <emglab_version>0.01</emglab_version>
  2) Root Element <emglab_spike_header>
    <time></time>
    <unit></unit>
    <emglab_spike_header>
    <!-- time unit -->
    <emglab_spike_events>
      0 1
      0.1111111111111111 2
      2 1
      3.141592653589793 0
      4 1
      5 2
      6 2
      7 1
    </emglab_spike_events>
  </emglab_annotation_file>
  6) Spike events: one time, spike ID per row.
  5) Spike header: Indicates each spike specifies time, spike ID
    
```

EAF: Complete Example

- Example includes all optional sections
 - All required, optional structure fields
 - Freeform variables
 - Includes structures, cell arrays, strings



Comparing Annotations

Annotation Comparison Methods

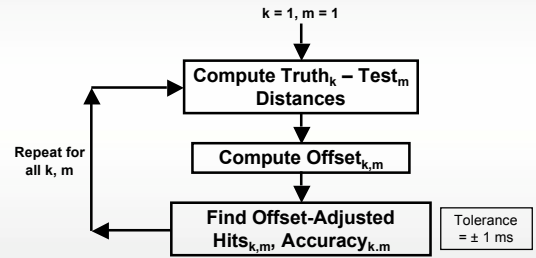
- Compare “truth” to “test”
 - “Truth” is known correct → Find errors
- or
- Compare “test1” vs. “test2” → Find differences
- **Issues:**
 - MU IDs vary file-to-file
 - Timing offset file-to-file
 - Time fiducials vary with method (e.g., peak, center of mass)
 - MUs have different SNRs
 - Compute results for each MU

Annotation Comparison: Resources

- Ambulatory electrocardiographs. American National Standard, ANSI/AAMI EC38:1998. Assoc Advance Med Instru, 1999.
- Farina D, Colombo R, Merletti R, Olsen HB. Evaluation of intramuscular EMG signal decomposition algorithms. *J Electromyography Kinesiol* 11:175-187, 2001.
 - Largely the method used in EMGLab
- Stashuk DW, Farina D, Søgaard K. Decomposition of Intramuscular EMG Signals. In: *Electromyography: Physiology, Engineering and Noninvasive Applications*, R Merletti and PA Parker eds., John Wiley and Sons, Inc., New Jersey, 47-80, 2004.
- Carey RM, Clancy EA. EMG decomposition annotation comparison method. *Proc IEEE 31st Ann Northeast Bioeng Conf*, IEEE, 100-101, 2005.

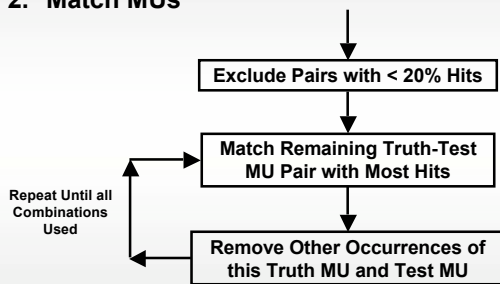
Annot Compare: Step 1 of 4

1. Compute match statistics



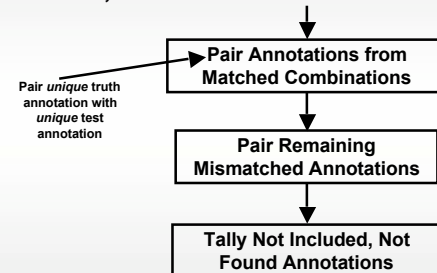
Annot Compare: Step 2 of 4

2. Match MUs



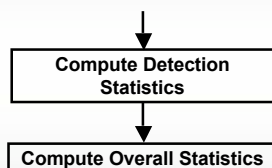
Annot Compare: Step 3 of 4

3. Pair; Form Confusion Matrix



Annot Compare: Step 4 of 4

• Compute Detection and Classification Performance Metrics



Detection Definitions

Parameter Name	General Definition	Interpretation for Detection Performance	Interpretation for Overall Performance
True Positive (TP)	A correctly detected spike.	A truth spike that is paired with <i>any</i> test spike.	A truth spike that is paired with a test spike from the matching test identity.
False Negative (FN)	A missed (erroneously rejected) spike.	A truth spike that is <i>not</i> paired with <i>any</i> test spike.	A truth spike that is <i>not</i> paired with a test spike from the matching test identity. Either it is incorrectly paired or it is not paired at all.
False Positive (FP)	An erroneously detected non-spike.	A test spike that is <i>not</i> paired with <i>any</i> truth spike.	A test spike that is <i>not</i> paired with any truth spike.

Intellectual Property (IP) Issues

What is Protected (U.S.A.)?

- **Protected:**
 - Software source code
 - Database *mode!*
- **Not protected:**
 - Raw physiologic data (not a “creative work”)
 - See [Science Commons FAQs](#)
 - Still, best to extend re-use rights with the data
- **Questionable:**
 - Annotation files: Likely protected
 - Especially if manually edited (creative work)
 - Data header files: ??
- European Union has additional specific database protections

Copyright License Resources

- **GNU Public Licenses**
 - Free Software Foundation (Richard Stallman)
 - <http://www.fsf.org>
- **Creative Commons Licenses**
 - Creative Commons Org. (Lawrence Lessig, Stanford Law)
 - <http://www.creativecommons.org>
- **Science Commons**
 - <http://www.sciencecommons.org>
- **Aladdin Free Public License**
 - Aladdin Enterprises
 - <http://www.cs.wisc.edu/~ghost/doc/cvs/Public.htm>



Licensing Issues

- **Open Source** ✓
 - Source code provided with distribution
 - Usually no cost
 - Source modification, sharing usually allowed
- **Source Modification** ✓
 - Right to modify, share source
- **Inherited Rights** ✓
 - If share software, must *give* same rights
 - Preserves openness, software freedom
- **Attribution** ✓
 - Credit authors/developers
- **Sell Program Copies / Commercial Use** X
 - Can combine, sell with proprietary code

Commercial Use

- **GPL**
 - Can sell with proprietary software **ONLY IF** “unbundled”
 - “Lesser” GPL: Can link to non-free software
 - LGPL usually applied to library software
- **Aladdin**
 - Cannot sell
 - Cannot ship with sold software
 - Can redistribute without cost

License Comparison

	GNU	Creative Commons	Aladdin
Inherited Rights, Attribution ?	Yes	Yes	Yes
Commercial Use ?	Limited	Yes/No (Selectable)	**NO**
Modify Source ?	Yes	Yes/No (Selectable)	Yes

↑ ↑ ↑
Not recommended for software

The Future of Open-Source EMG Software

...

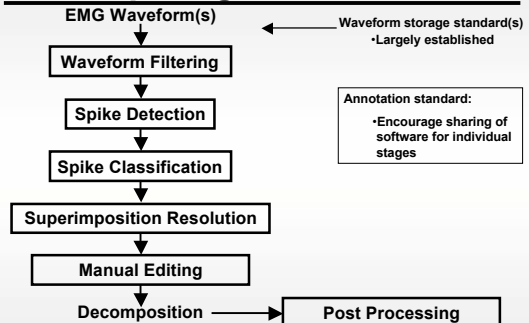
EMGlab Website Tools

- **Data Format: WaveForm DataBase (WFDB) standard** (<http://www.physionet.org>)
- **Automated decomposition (MATLAB code)**
 - McGill "autodecomp" (with superimposition resolution)
 - Florestal/Montreal decomp algorithm
- **EMGlab annotation format**
 - Definition, documentation, MATLAB code
- **Annotation comparison**
 - Definition, MATLAB code, viewer
- **Manual reviewing/editing tool**
- **EMGlab MATLAB software**
 - MATLAB GUI
 - Links tools for complete decomposition (or any piece)
 - <http://www.emglab.net>

EMGlab Extensions/Development

- **Multiple-channel, single site**
 - Some aspects already available
- **Coordinated modification by other users**
 - Mechanisms to share enhancements
 - Additional features of use to others
 - Extensions

Decomp Stages: "Traditional"



Other Possible Shared Resources

- **EMG simulators** (Outputs in EMGlab formats)
 - Based on electrophysiologic model
 - Based on physiologic templates
 - Hamilton-Wright, Stashuk simulator recently added!
- **Annotation post-processing** (Inputs in EAF format)
 - Firing rate, synchronization
- **DeCompetition(s)**
- **Teaching modules**
 - Self-study based on Merletti-Parker book