Virtual Therapy in Head&Neck cancer
Data collection and sharing for articulatory synthesis of patient speech

Rob van Son

Netherlands Cancer Institute, Amsterdam, The Netherlands
R.v.Son@nki.nl

15 November 2017
TAPAS: Training Network on Automatic Processing of Pathological Speech

EU promotes Open Data and Open Science

- 15 Early Stage Researchers working in 9 countries
- TAPAS will collect unique data on speakers and pathological speech
- Speakers are “vulnerable”: patients and children
- Most stringent requirements on privacy protection (GDPR&CTR)
- But we still want to share
Collections of Patient Speech and PROs

Current practices regarding speech data

NKI-AVL (hospital), ACLC (university)
Data of Head & Neck tumor patients at NKI-AVL

### Protocols: Speech and Patient Reported Outcomes (PROs\(^1\))

- **Data pre- and post-treatment (upto 12mnd)**
- **Oral Cavity protocol (\(~95\))**
  - Sustained a:-u:-i:; ei-au-œy; word-list (36w); story\(^\dagger\) (75w); pa-ta-ka
  - PROs: SHI, SWAL QoL, EORTC QLQ-H&N35
- **Larynx protocol (\(~150\))**
  - a: Longest/high/low/loud/soft/sweep; story\(^\dagger\)
  - PROs: LASA, EAT-10, VHI, EORTC QLQ-C30/H&N35
- **Tracheolaryngectomy (TLE) protocol (\(~25\))**
  - a: Longest/high/low/loud/soft/sweep; story\(^\dagger\); 3 voiced sentences\(^*\)
  - PROs: EQ-5D-5L, SOAL, VHI-10,

---

\(^1\)PROs are questionnaires asking about health and QoL
Secondary use of patient data

Currently *opt-out*

- Health care data
- Restricted use (*severely restricted under opt-out*)
- IRB approval needed for *each* project
- Can we create a speech “Biobank” from secondary use data? *Probably not*

Switch to *Informed Consent* (opt-in)

- All patients asked to consent
- Look for broad consent
  *Unclear whether broad consent is possible*
pre-2000 longitudinal research of speech development at the ACLC\textsuperscript{2}

Recordings of mother-child interactions

- Video of 12 Cleft Palate + 6 normal babies [1]
- Audio of 5 hearing impaired and 5 normal hearing babies [2, 3]
- Recorded 1 × month during first 2 years
- Informed consent of parents for research
- Video material digitized, text files are now being digitized
- Intention to make them available to researchers
- Access policies not yet known

\textsuperscript{2}Amsterdam Centre for Language and Communication, Fac. Hum., Univ. of Amsterdam
Virtual Therapy

Predicting and synthesizing plausible speech examples after oral cancer treatment
Demonstrating functional outcome of therapy

AREA 1: IMAGING
- Anatomy
- Integration
- Dynamics

AREA 2: MODELLING
- Simulation
- Function
- Digital Patient

AREA 3: USER INTERFACE
- User Interface

Data for articulatory synthesis
Virtual Therapy

Data collection

Digital Patient

1. Anatomy & Physiology
   MRI, DTI-MRI, Shear-Wave US elasticity
   3D Photo of tongue shape and resection

2. Dynamics
   dynamic MRI, Tongue mobility & strength,
   3D video, sEMG

3. Integration
   ArtiSynth bio-mechanical model

4. Simulation
   Forward and inverse model training

5. Function (Speech)
   Real Speech & Articulatory synthesis

20 patients + 10 healthy

- 10 small tumors in oral tongue (surgery)
- 10 larger tumors in base of tongue (radio-therapy)
Data Sharing

What can be shared, when, and with whom?
Confusion about *Informed Consent* (IC) in Open Data/Science

- IC determines what can be done with the data, i.e., cover all uses
- IC covered by GDPR and CTR³ [4, 5]
  - GDPR: Consent must be specific [6, 7]
  - CTR: One-stop informed consent possible [8, 9]
- EU vs. National rules on health data and consent (CTR, [7])
- What health data fall under the research derogation of GDPR, if any?
- What research is “in the public interest”?
- Open data is international, the GDPR restricts cross-border exchange

³Clinical Trials Regulation (536/2014)
Data Sharing

Consensus: Privacy by design

Demands on shared data (under the GDRP)

- Data minimization
  - what is not there, cannot be exposed
  - Coarse-graining: age-brackets, truncate zip codes, etc.
  - Strip metadata from images, movies, MRI
  - Censor bars in pictures, movies, MRI

- Anonymization
  - if data is useful, it is not anonymous

- Pseudonymization
  - is effort needed to re-identify relevant?

- Encryption

- Security, computer and otherwise

- Data transfer agreements, NDA’s, Promise of Confidentiality

⇒ Take the analysis to the data

privacy-preserving platform
Take the analysis to the data

On-demand analysis on a privacy-preserving platform

- Users do not see micro-data or individual records
- Users only see the outcome of the certified analysis
- Platform supplies tools: e.g., R, Bob and Kaldi \cite{10, 11}
- Platform enforces access rights, audit trails, and security
- Only a single DTA contract between data-owner and platform owner
The BEAT platform

Biometrics Evaluation and Testing (BEAT)

- EU 7th framework program
- Part of European computing e-infrastructure for Open Science
- Solution for open access, scientific information sharing and re-use
- Sharing access to data and source code
- Protecting privacy and confidentiality
- Data from different experiments can be easily compared and searched
- Challenges and education
- Attestation mechanism for reports

⇒ Chosen as the data platform for TAPAS

BEAT overview

Figure 1: An overview of the BEAT platform applications and their interaction. Users use the web front-end to run experiments, search and combine results. A back-end handles the execution of experiments on dedicated hardware.
BEAT toolchain block

Individual Blocks are strung together into tool-chains/experiments
Data Sharing

BEAT experiment configurator

Figure 3: The BEAT platform experiment configurator allows the user to easily associate databases, algorithms and analyzers together to create the desired setup. As the user chooses components for the toolchain, choices of further components are restricted respecting data format compatibility between the blocks.
BEAT automatically generated tables
Conclusions

TAPAS: Sharing data

- Not clear what will be allowed under the GDPR&CTR
- “International” sharing could be a problem (outside EU)
- In the mean time: use Privacy Preserving Platform

⇒ TAPAS will use BEAT platform
Thank You!
More information I


More information II


More information III


This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

©2017 R.J.J.H. van Son