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

NETHERLANDS CANCER INSTITUTE
ANTONI VAN LEEUWENHOEK

UNIVERSITY OF AMSTERDAM
Amsterdam Center for Language and Communication

Introduction

NETHERLANDS CANCER INSTITUTE
ANTONI VAN LEEUWENHOEK

UNIVERSITY OF AMSTERDAM
Amsterdam Center for Language and Communication

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 aː-uː-iː; ei-ɑu-œy; word-list (36w); story† (75w); pa-ta-ka
  - PROs: SHI, SWAL QoL, EORTC QLQ-H&N35
- Larynx protocol (∼150)
  - aː Longest/high/low/loud/soft/sweep; story†
  - PROs: LASA, EAT-10, VHI, EORTC QLQ-C30/H&N35
- Tracheolaryngectomy (TLE) protocol (∼25)
  - aː Longest/high/low/loud/soft/sweep; story†; 3 voiced sentences*
  - PROs: EQ-5D-5L, SOAL, VHI-10,

[1] PROs are questionaires 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[2]

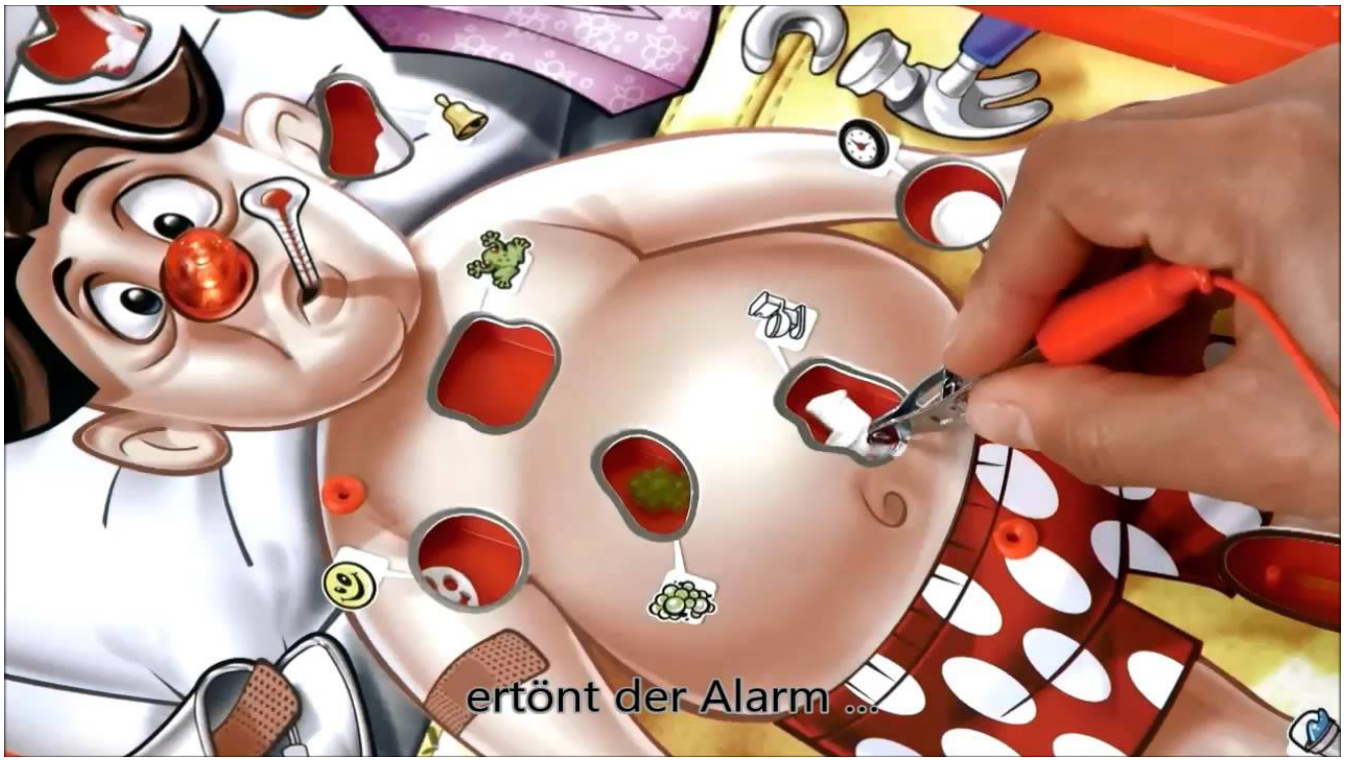
Recordings of mother-child interactions

- Video of 12 Cleft Palate + 6 normal babies [1]
- Audio of 5 hearing impaired and 5 normal hearing babies [2, 3]
- Recorded 1 x 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

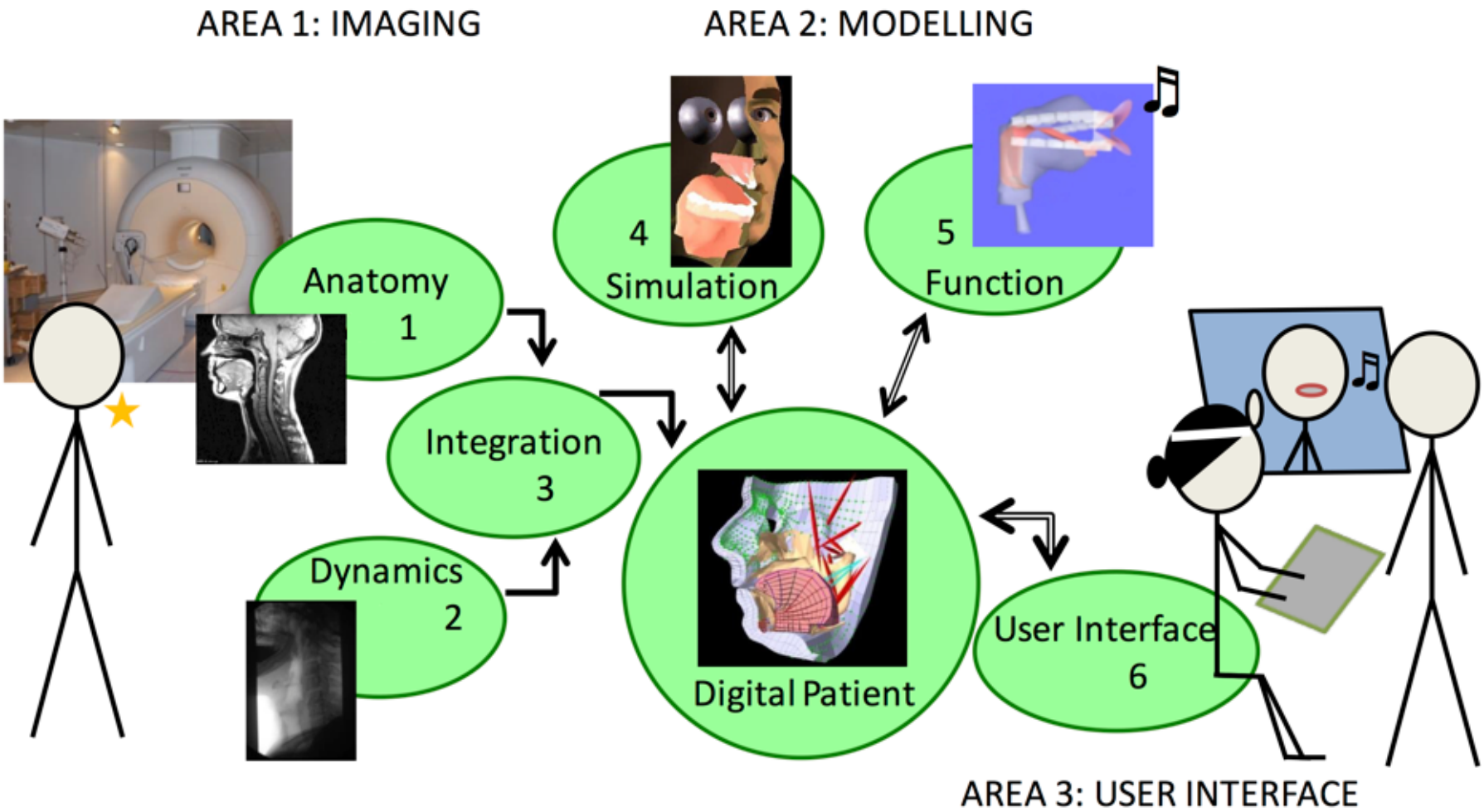
[2] Amsterdam Centre for Language and Communication, Fac. Hum., Univ. of Amsterdam

Virtual Therapy

Predicting and synthesizing plausible speech examples after oral cancer treatment



Demonstrating functional outcome of therapy



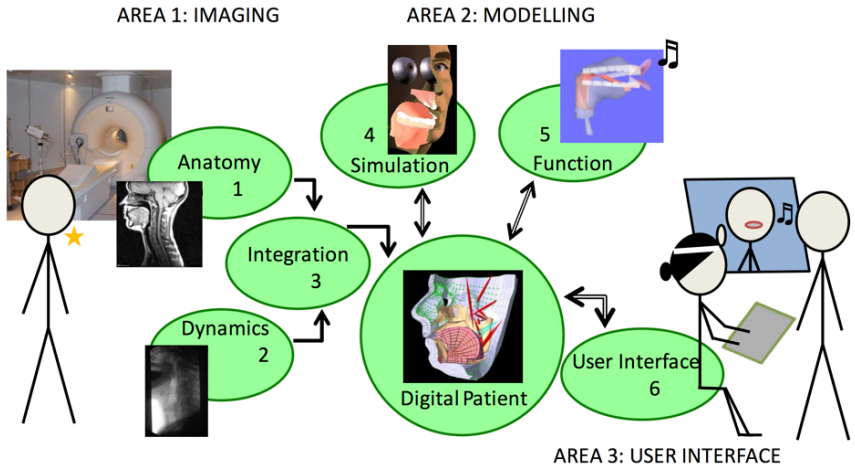
Data collection

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?

Talk to:
MEC, Privacy Officer, Biobank organizers, Privacy lawyers

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 CTR[3] [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

[3] Clinical Trials Regulation (536/2014)

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

$\Rightarrow$ Take the analysis to the data — *privacy-preserving platform*

Take the analysis to the data

On-demand analysis on a privacy-preserving platform

- User 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 [10, 11]
- Platform enforces access rights, audit trails, and security
- Only a single DTA contract between data-owner and platform owner

The BEAT platform [13]

Biometrics Evaluation and Testing (BEAT) [12]

- EU 7^{th} 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

$\Rightarrow$ Chosen as the data platform for TAPAS

Anjos et al. (2017). *BEAT: An Open-Source Web-Based Open-Science Platform*. arXiv preprint arXiv:1704.02319.

BEAT overview

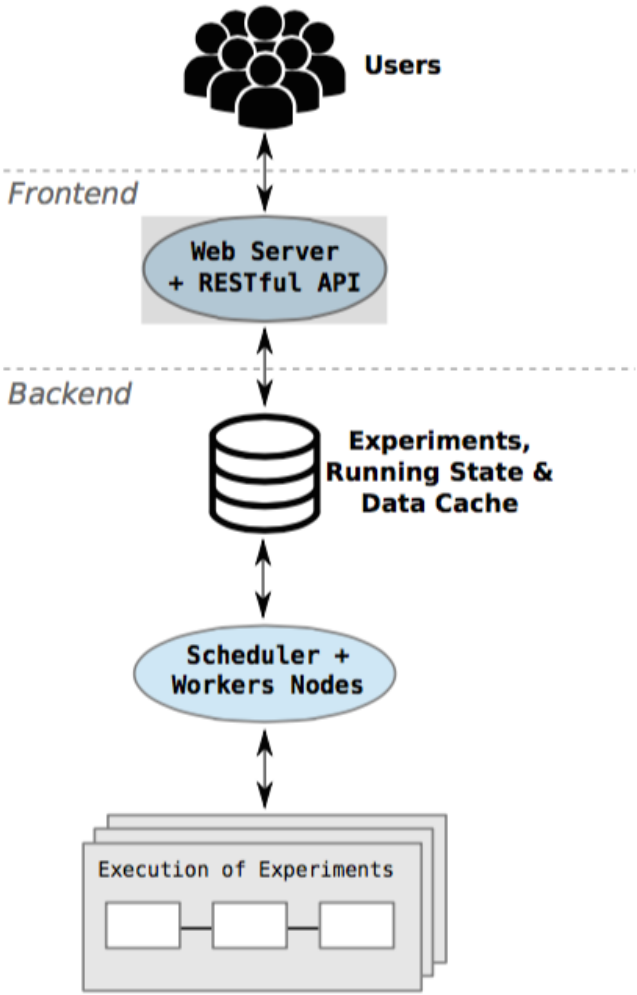
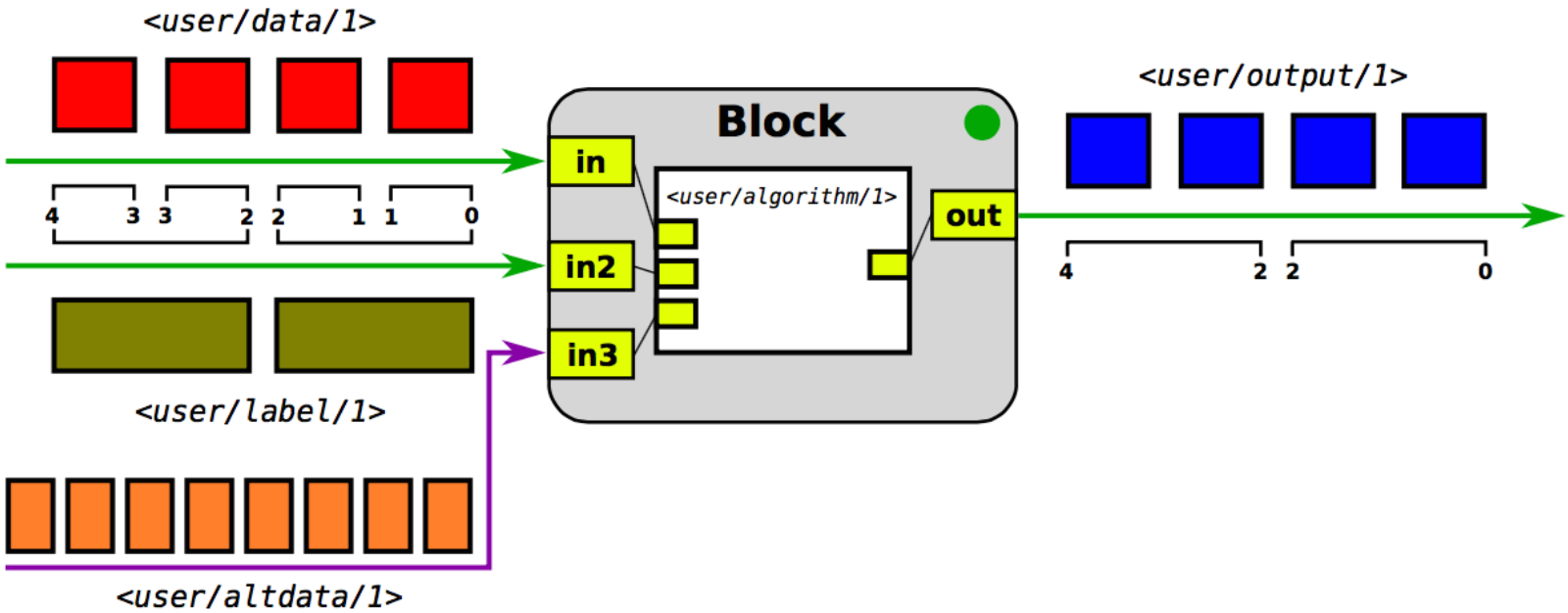


Figure 1: An overview of the BEAT platform applications and their interaction. Users use the web front-end to run experiments, search and combine results. A back-end handles the execution of experiments on dedicated hardware.

BEAT toolchain block



Individual Blocks are strung together into tool-chains/experiments

BEAT experiment configurator

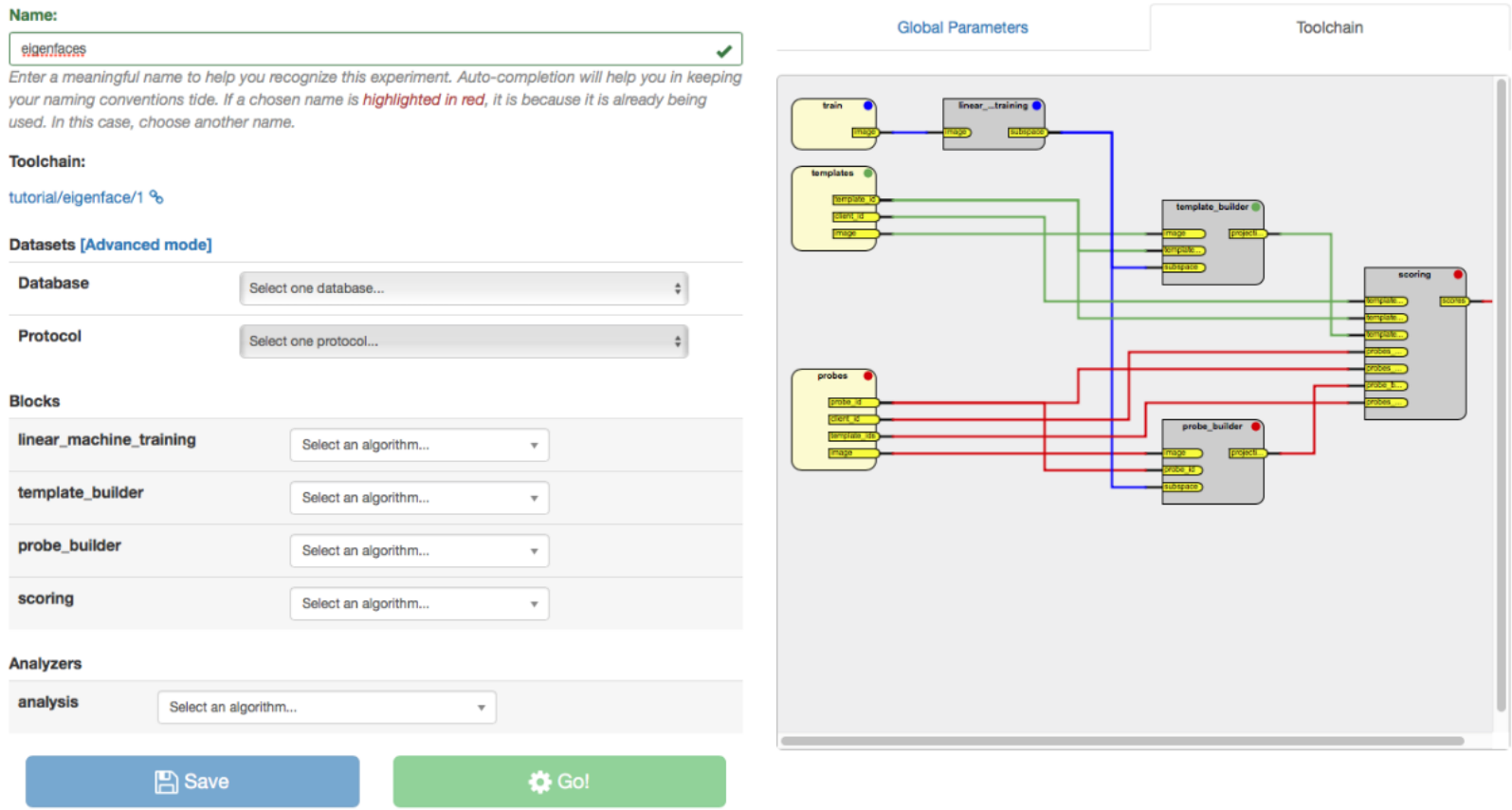


Figure 3: The BEAT platform experiment configurator allows the user to easily associate databases, algorithms and analyzers together to create the desired setup. As the user chooses components for the toolchain, choices of further components are restricted respecting data format compatibility between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

BEAT automatically generated tables

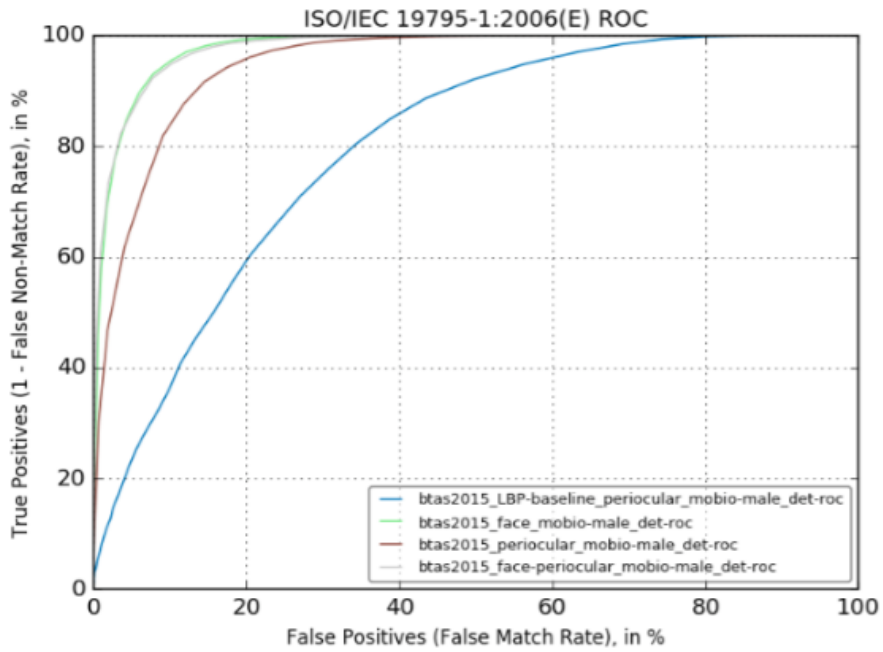
Add to report

roc_test × frr_test × eer × hter × roc_dev × far_test ×

Updated	Name	frr_test	eer	hter	far_test
Sept. 1, 2015	tpereira/tpereira/full_isv/2/btas2015_face_mobio-male_det	0.10802	0.0301587	0.0739247	0.0398293
Sept. 1, 2015	tpereira/tpereira/full_isv_multi/2/btas2015_face-periocular_mobio-male_det	0.0977444	0.0384921	0.0756452	0.053546
Sept. 1, 2015	tpereira/tpereira/full_isv/2/btas2015_periocular_mobio-male_det	0.132832	0.0845238	0.116589	0.100345
Sept. 1, 2015	tpereira/tutorial/full_lbphs/2/btas2015_LBP-baseline_periocular_mobio-male_det	0.294987	0.242391	0.274826	0.254664

roc_test

Plotter parameter: plot/isoroc/1 Expand



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

$\Rightarrow$ TAPAS will use BEAT platform

Thank You!

?

More information I

[1] F. J. Koopmans-van Beinum and J. M. van der Stelt, "Early stages in the development of speech movements," in *Precursors of early speech*, pp. 37–50, Springer, 1986.

[2] C. Clement, E. den Os, and F. Koopmans-van Beinum, "The development of vocalizations of deaf and normally hearing infants," *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, vol. 18, pp. 65–76, Available at `http://www.fon.hum.uva.nl/archive/1994/Proc18ClementOsBeinum.pdf`, 1994.

[3] J. M. van der Stelt, T. G. Wempe, and L. C. W. Pols, "Comparing deaf and hearing dutch infants: Changes in the vowel space in the first 2 years," *Clinical Linguistics & Phonetics*, vol. 22, no. 10-11, pp. 835–844, doi:10.1080/02699200802175842, 2008. PMID: 18608233.

[4] European Parliament and the Council of the EU, "Clinical trials regulation 536/2014." `https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2014_536/reg_2014_536_en.pdf`, 2014.

[5] C. K. Schneider *et al.*, "Deciphering the eu clinical trials regulation," *Nature biotechnology*, vol. 34, no. 3, p. 231, 2016.

[6] European Union, "General Data Protection Regulation, GDPR." `https://gdpr-info.eu/`, 2016.

More information II

[7] G. Chassang, "The impact of the eu general data protection regulation on scientific research," *ecancermedicalscience*, vol. 11, pp. 709, doi:10.3332/ecancer.2017.709, 2017.

[8] A. Mende, M. Frech, and C. Riedel, "Principles of the eu clinical trials regulation no 536/2014: What will change?," *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 2017.

[9] C. Dittrich, A. Negrouk, and P. G. a. Casali, "An ESMO-EORTC position paper on the EU clinical trials regulation and EMA's transparency policy: making european research more competitive again," *Annals of Oncology*, vol. 26, no. 5, pp. 829–832, doi:10.1093/annonc/mdv154, 2015.

[10] M. Cernak, A. Komaty, A. Mohammadi, A. Anjos, and S. Marcel, "Bob speaks kaldi," in *Proc. of Interspeech*, no. EPFL-CONF-229211, 2017.

[11] A. Gaye, Y. Marcon, J. Isaeva, P. LaFlamme, A. Turner, ..., and P. R. Burton, "DataSHIELD: taking the analysis to the data, not the data to the analysis," *International Journal of Epidemiology*, vol. 43, no. 6, pp. 1929–1944, doi:10.1093/ije/dyu188, 2014.

[12] Idiap Research Institute, "Biometrics Evaluation and Testing (BEAT) ." `https://www.beat-eu.org/`, 2017.

More information III

[13] A. Anjos, L. El-Shafey, and S. Marcel, "Beat: An open-source web-based open-science platform," *arXiv preprint arXiv:1704.02319*, 2017.