Some applications of speech processing in medical settings, from communication to diagnosis

Saturnino Luz

CBMS 2018, Karlstad, Sweden, June 2018

Usher Institute
Population Health Sciences & Informatics

THE UNIVERSITY of EDINBURGH
Edinburgh Medical School

Molecular, Genetic and Population Health Sciences

Usher Institute, Edinburgh Medical School

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Speech, Language and Communication in Medicine

- Medical teams
- Clinician-patient communication
- Speech rhythm as a 'biomarker'?

Multidisciplinary Medical Team Meetings (MDTMs)

Bridget Kane and Saturnino Luz. Multidisciplinary medical team meetings: An analysis of collaborative working with special attention to timing and teleconferencing. *Computer Supported Cooperative Work (CSCW)*, 15(5):501–535, 2006. doi: 10.1007/s10606-006-9035-y

Bridget Kane and Saturnino Luz. Achieving diagnosis by consensus. *Computer Supported Cooperative Work (CSCW)*, 18(4):357–391, 2009. doi: 10.1007/s10606-009-9094-y

MDTM analysis

- Initial study: dependability issues related to MDTMs supported by teleconference (TeleSynergy^{tm})
- Mostly qualitative/descriptive work but with quantitative elements.
- Notably, anaysis of conversational turns (vocalisations) in fully co-located meetings and teleconference
- Analysis inspired by work on remote communication, by Sellen and others.

Abigail J. Sellen. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10(4):401–444, 1995. doi: 10.1207/s15327051hci1004_2. URL http://www.leaonline.com/doi/abs/10.1207/s15327051hci1004_2

MDTM analysis

- Initial study: dependability issues related to MDTMs supported by teleconference (TeleSynergy^{tm})
- Mostly qualitative/descriptive work but with quantitative elements.
- Notably, anaysis of conversational turns (vocalisations) in fully co-located meetings and teleconference
- Analysis inspired by work on remote communication, by Sellen and others.

Abigail J. Sellen. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10(4):401–444, 1995.
doi: 10.1207/s15327051hci1004_2.
URL http://www.leaonline.com/doi/abs/10.1207/s15327051hci1004_2

Differences in remote conversations

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Table: **Average number and duration of individual turns, people participating, informal exchanges, group turns, silences and turn variability (H) per case.**

Parameter	t-test	Tele-conference	Co-located
Time spent per case	$p < 0.05$	x	$0.4x$
Number of turns/min. per case	$p < 0.05$	5.3	8.6
Turn Duration (in seconds)	$p < 0.01$	13.14	7.42
No of active participants	$p < 0.01$	8.62	7.00
% informal conversation	$p < 0.05$	0.13	2.26
Turns per person per case	not sig.	4.06	3.48
Group turn duration per case	not sig.	2.17	1.66
No. Group turns/min. per case	not sig.	0.23	0.372
% silence per case	not sig.	3.28	4.29
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76

Differences in remote conversations

Table: **Average number and duration of individual turns, people participating, informal exchanges, group turns, silences and turn variability (H) per case.**

Parameter	t-test	Tele-conference	Co-located
Time spent per case	$p < 0.05$	x	$0.4x$
Number of turns/min. per case	$p < 0.05$	5.3	8.6
Turn Duration (in seconds)	$p < 0.01$	13.14	7.42
No of active participants	$p < 0.01$	8.62	7.00
% informal conversation	$p < 0.05$	0.13	2.26
Turns per person per case	not sig.	4.06	3.48
Group turn duration per case	not sig.	2.17	1.66
No. Group turns/min. per case	not sig.	0.23	0.372
% silence per case	not sig.	3.28	4.29
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76

Differences in remote conversations

Table: **Average number and duration of individual turns, people participating, informal exchanges, group turns, silences and turn variability (H) per case.**

Parameter	t-test	Tele-conference	Co-located
Time spent per case	$p < 0.05$	x	$0.4x$
Number of turns/min. per case	$p < 0.05$	5.3	8.6
Turn Duration (in seconds)	$p < 0.01$	13.14	7.42
No of active participants	$p < 0.01$	8.62	7.00
% informal conversation	$p < 0.05$	0.13	2.26
Turns per person per case	not sig.	4.06	3.48
Group turn duration per case	not sig.	2.17	1.66
No. Group turns/min. per case	not sig.	0.23	0.372
% silence per case	not sig.	3.28	4.29
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76

Differences in remote conversations

Table: **Average number and duration of individual turns, people participating, informal exchanges, group turns, silences and turn variability (H) per case.**

Parameter	t-test	Tele-conference	Co-located
Time spent per case	$p < 0.05$	x	$0.4x$
Number of turns/min. per case	$p < 0.05$	5.3	8.6
Turn Duration (in seconds)	$p < 0.01$	13.14	7.42
No of active participants	$p < 0.01$	8.62	7.00
% informal conversation	$p < 0.05$	0.13	2.26
Turns per person per case	not sig.	4.06	3.48
Group turn duration per case	not sig.	2.17	1.66
No. Group turns/min. per case	not sig.	0.23	0.372
% silence per case	not sig.	3.28	4.29
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76

Differences in remote conversations

Table: **Average number and duration of individual turns, people participating, informal exchanges, group turns, silences and turn variability (H) per case.**

Parameter	t-test	Tele-conference	Co-located
Time spent per case	$p < 0.05$	x	$0.4x$
Number of turns/min. per case	$p < 0.05$	5.3	8.6
Turn Duration (in seconds)	$p < 0.01$	13.14	7.42
No of active participants	$p < 0.01$	8.62	7.00
% informal conversation	$p < 0.05$	0.13	2.26
Turns per person per case	not sig.	4.06	3.48
Group turn duration per case	not sig.	2.17	1.66
No. Group turns/min. per case	not sig.	0.23	0.372
% silence per case	not sig.	3.28	4.29
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76

The effect of distance on the participant's requirements

Table: **Participant's requirements expressed *prior* to teleconferencing experience and revised requirement for video *after* 8 months experience.**

Requirement	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
See all Radiology	74%	11%	16%	0	0
See all Pathology	58%	26%	11%	5%	0
Hear all Participants	21%	32%	42%	5%	0
See all Participants '04	5%	21%	53%	21%	0
See all Participants '05	19%	50%	19%	8%	4

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

The effect of distance on the participant's requirements

Table: **Participant's requirements expressed *prior* to teleconferencing experience and revised requirement for video *after* 8 months experience.**

Requirement	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
See all Radiology	74%	11%	16%	0	0
See all Pathology	58%	26%	11%	5%	0
Hear all Participants	21%	32%	42%	5%	0
See all Participants '04	5%	21%	53%	21%	0
See all Participants '05	19%	50%	19%	8%	4

The effect of distance on the participant's requirements

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Table: **Participant's requirements expressed *prior* to teleconferencing experience and revised requirement for video *after* 8 months experience.**

Requirement	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
See all Radiology	74%	11%	16%	0	0
See all Pathology	58%	26%	11%	5%	0
Hear all Participants	21%	32%	42%	5%	0
See all Participants '04	5%	21%	53%	21%	0
See all Participants '05	19%	50%	19%	8%	4

The effect of distance on the display of information

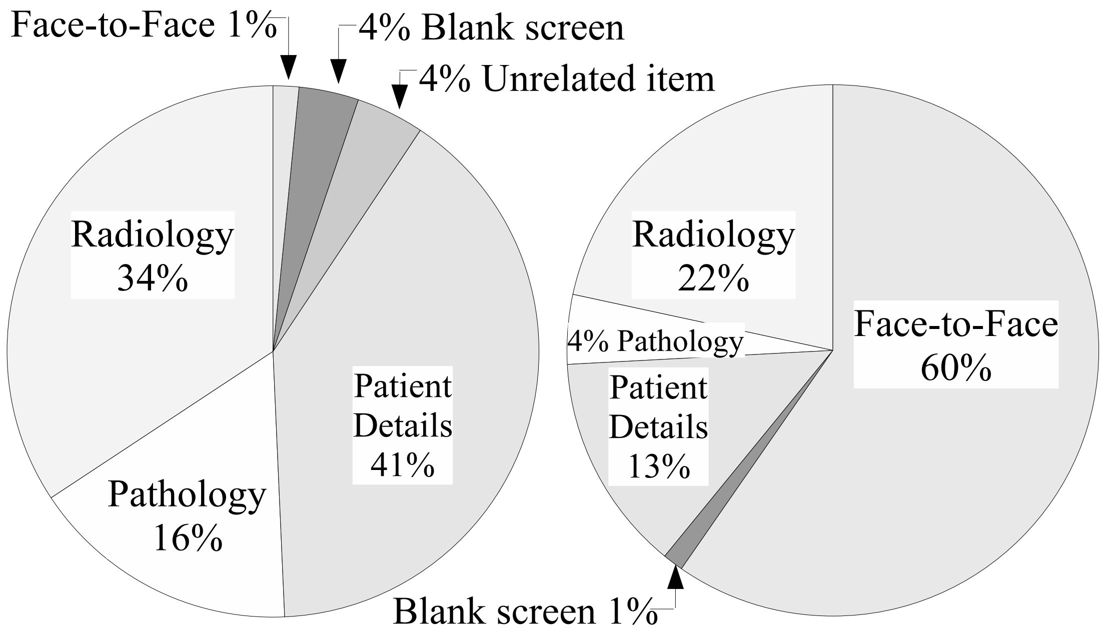


Figure: Display use at co-located (left) and teleconference (right) meetings

Beyond quantitative analysis

Further issues:

- Time, Schedules, Place,
- Other Functions (socio-emotional, educational),
- Volume of Work: Preparation, Work Generated, Feedback
- Decision-making: Evidence Base, Rationale, Record-keeping, Validation,
- Record keeping, ...
  - A role for speech processing?

Parameter	t-test	Tele-conference	Co-located
...			
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76
...			

Bridget Kane and Saturnino Luz. "Do no harm": Fortifying MDT collaboration in changing technological times. *International Journal of Medical Informatics*, 82(7):613–625, July 2013. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2013.03.003

Bridget T. Kane, Pieter J. Toussaint, and Saturnino Luz. Shared decision making needs a communication record. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 79–90, New York, NY, USA, 2013. ACM. doi: 10.1145/2441776.2441788

Beyond quantitative analysis

Further issues:

- Time, Schedules, Place,
- Other Functions (socio-emotional, educational),
- Volume of Work: Preparation, Work Generated, Feedback
- Decision-making: Evidence Base, Rationale, Record-keeping, Validation,
- Record keeping, ...
  - A role for speech processing?

Parameter	t-test	Tele-conference	Co-located
...			
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76
...			

Bridget Kane and Saturnino Luz. "Do no harm": Fortifying MDT collaboration in changing technological times. *International Journal of Medical Informatics*, 82(7):613–625, July 2013. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2013.03.003

Bridget T. Kane, Pieter J. Toussaint, and Saturnino Luz. Shared decision making needs a communication record. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 79–90, New York, NY, USA, 2013. ACM. doi: 10.1145/2441776.2441788

Beyond quantitative analysis

Further issues:

- Time, Schedules, Place,
- Other Functions (socio-emotional, educational),
- Volume of Work: Preparation, Work Generated, Feedback
- Decision-making: Evidence Base, Rationale, Record-keeping, Validation,
- Record keeping, ...
  - A role for speech processing?

Parameter	t-test	Tele-conference	Co-located
...			
Turn distribution (entropy)	not sig.	2.422	2.250
Turn distribution (η)	not sig.	0.74	0.76
...			

Bridget Kane and Saturnino Luz. "Do no harm": Fortifying MDT collaboration in changing technological times. *International Journal of Medical Informatics*, 82(7):613–625, July 2013. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2013.03.003

Bridget T. Kane, Pieter J. Toussaint, and Saturnino Luz. Shared decision making needs a communication record. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 79–90, New York, NY, USA, 2013. ACM. doi: 10.1145/2441776.2441788

Electronic record-keeping

“By the time you have moved your hand from the keyboard, reached for the mouse, pointed, and click or whatever ... youve lost too much time ... and the discussion has moved on”

Opportunities

- Technology to support
  - Preparation: information retrieval, text analysis
  - Collaboration: Retrieval, image processing, NLP
  - Sharing: Data / Information Capture
  - Re-use of information:
    - Education: Informative Patient Case Discussions
    - Reviewing:
      Did X say anything about option Y?
      Was TB considered when making the decision?

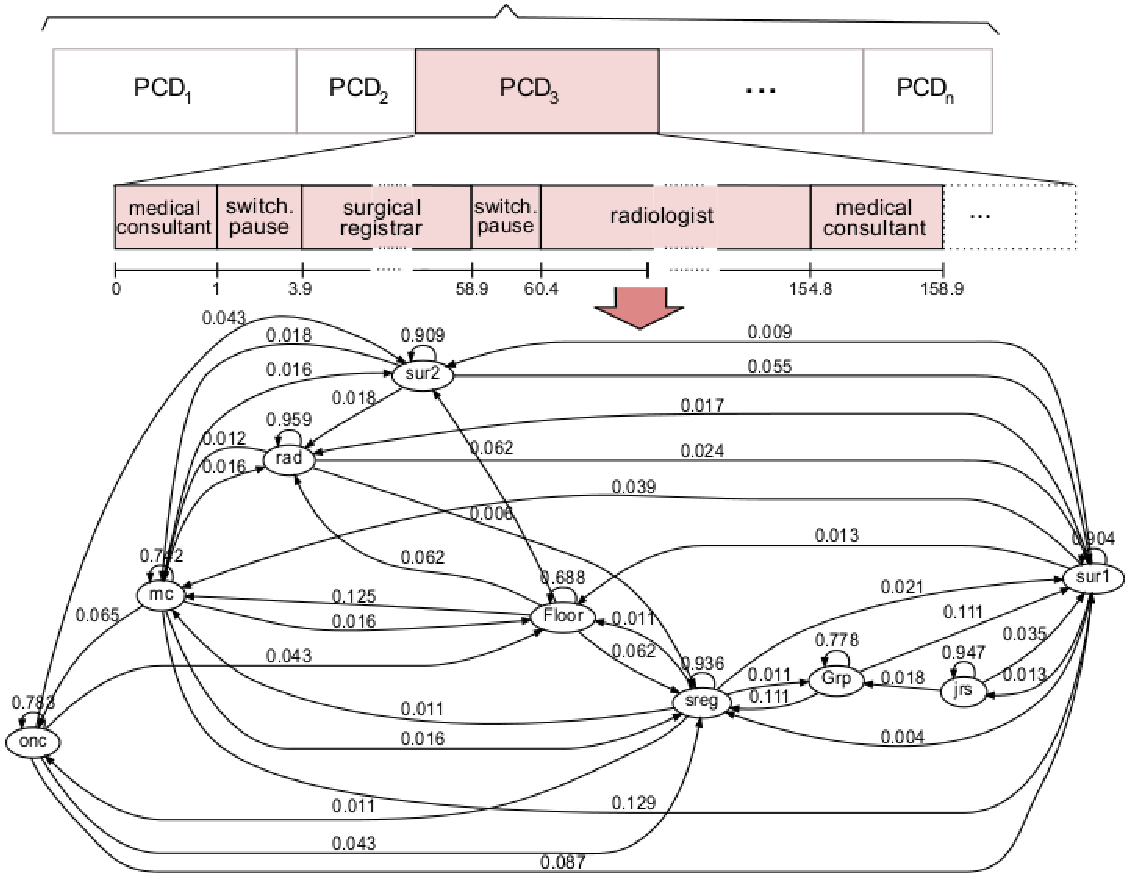
- Need for new methods to capture and transform information.

- The case of Patient Case Discussions (PCD)

Saturnino Luz and Bridget Kane. Perspectives on intelligent systems support for multidisciplinary medical teams.
In *AAAI Spring Symposium Series*, pages 272–275, 2016.
URL http://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12741/11964

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

The stability of vocalisation patterns



THE UNIVERSITY of EDINBURGH

P.T. Brady. A statistical analysis of on-off patterns in 16 conversations. *The Bell System Technical Journal*, 47:73–91, 1968

James M. Jr. Dabbs and Barry Ruback. Dimensions of group process: Amount and structure of vocal interaction. *Advances in Experimental Social Psychology*, 20(123–169), 1987

Identifying patterns and structure

- Analysis of meetings:
  - Interesting theoretical domain: understanting the mechanisms underlying multi-party dialogue and collaboration;
  - Long, multidisplinary research tradition e.g. [Bales, 1950]
  - Compelling application domain: meetings are ubiquitous (and not always productive).
- Related research:
  - Focus on automatic analysis of recorded meetings & user interaction: “meeting browsers”
  - Large data collection and analysis projects (AMI/AMIDA, ICSI, CMU etc)
  - But data come mostly from artificial settings...

A meeting segmentation task

- Data set: recorded (audio and video) MDTMs
- Task: segmentation of MDTMs, i.e. detection of Patient Case Discussion (PCD) boundaries in the recording, cast as a categorisation problem.
  - similarity with topic segmentation.
- Data representation: content-free patterns of vocalisation and silence
- Categorisation method: Naïve Bayes (combination of multinomial and continuous models)
- Results: comparable to those obtained by “content-rich” approaches

MDTM data set

- For this task: 54 PCDs annotated for boundaries, vocalisations (individual and group), silences (Pause, SwitchingPause, GrpPause and GrpSwitchingPause), and specialist roles.

Participation by roles:

Role	Vocal. (%)
radiologist	24.9
surgeon	22.8
medical consultant	16.0
clinical oncologist	8.2
medical registrar	5.6
oncologist	5.5
surgical registrar	5.0
junior doctors	4.2
pathologist	2.7
nurse	0.5

Vocalisation events as topic shift predictors

Table: Mean durations in seconds (and standard deviations) of vocalisation and pauses on and near topic boundaries in MDTM and AMI meetings.

	Boundary	Non-boundary	t-test
AMI vocalisations	5.3 (8.2)	1.6 (3.5)	$p < .01$
AMI pauses	2.6 (4.9)	1.2 (2.8)	$p < .01$
AMI overlaps	0.4 (0.4)	0.3 (0.6)	$p < .01$
MDTM vocalisations	12.0 (15.5)	8.1 (14.7)	$p < .05$
MDTM pauses	9.7 (12.7)	8.2 (14.8)	$p < .05$

Hypotheses

H_1: vocalisation patterns alone can produce segmentation accuracy significantly higher than baseline segmentation accuracy for MDTMs

$H_{1.1}$: segmentation performance is better when distinctions between pause types are included ($V_i \in Roles \cup \{Floor\}$ Vs. $V_i \in Roles \cup \{Pause, SwPause, GrpPause, GrpSwPause\}$)

$H_{1.2}$: segmentation performance increases with longer 'vocalisation horizons'.

Representation of turns

- Vocalisation graphs assume conversation modelled as a Markov process
- For segmentation, we allow representation of longer chains of vocalisation
- Data set consists of time intervals to be classified as boundary/non-boundary instances:

$$s = (S_0, D_0, S_{-1}, D_{-1} \ldots, S_{-n}, D_{-n}, S_1, D_1 \ldots, S_n, D_n) \tag{1}$$

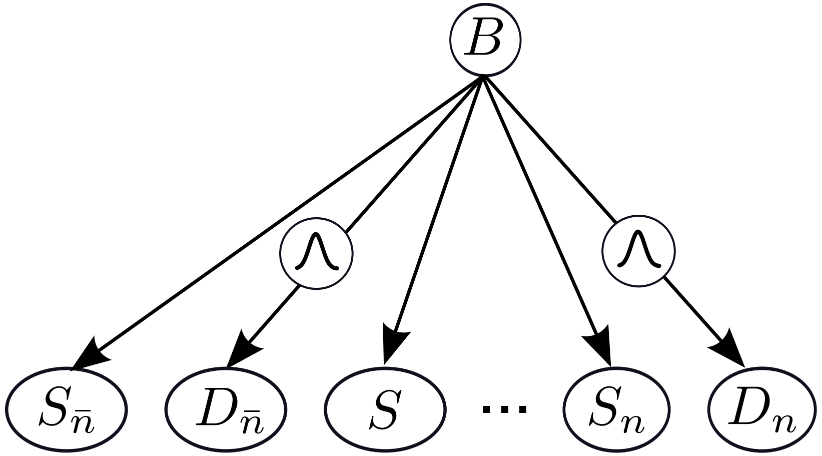
  - $S_0, \ldots, S_n$: speaker roles (sur, mc, rad, nurse etc) + Grp and Pause type
  - $D_0, \ldots, D_n$: duration of a speech (or silence) interval

Learning

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

- Naïve Bayes approach.
- Combine continuous and nominal features [John and Langley, 1995]:

$$
\begin{aligned}
P(b|V=v) &= P(b|S_{\bar{n}}=s_{\bar{n}}, D_{\bar{n}}=d_{\bar{n}}, \ldots, S=s, \ldots, D_n=d_n) \\
&\propto \prod_{i=-n}^{n} P(S_i|b)P(D_i|b)
\end{aligned}
$$

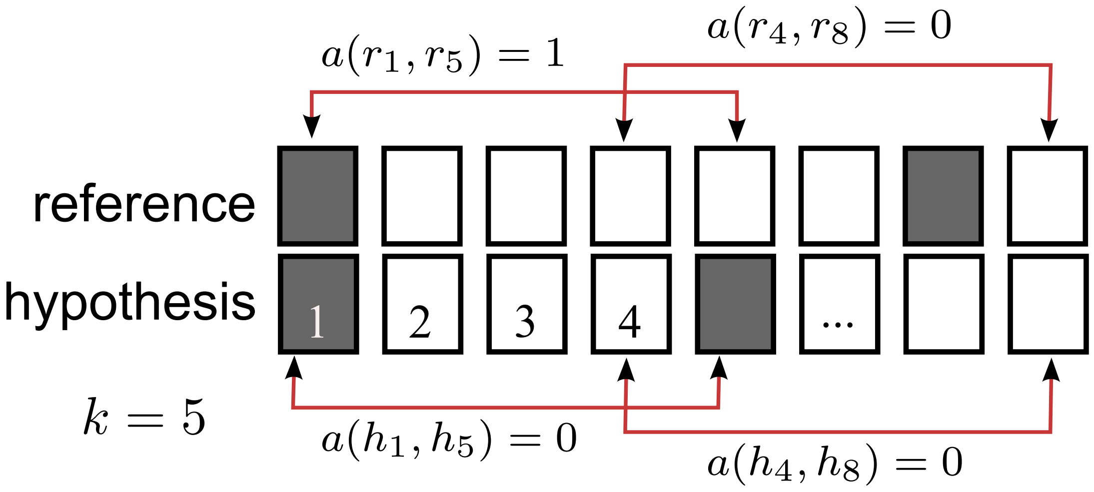


Hypotheses testing

- Cross validation experiments
- Comparison of two sequences:
  - reference: manually segmented, and
  - hypothesis: segmented according to a maximum *a posteriori* Naïve Bayes
- Varied:
  - Length of vocalisation horizon: $1 \leq n \leq 7$
  - Pause type discrimination: yes/no
- Standard IR metrics are inappropriate for evaluation:
  - highly imbalanced classes (only 3% are boundaries, so accuracy of 97% for trivial classifier)
  - F_+, π_+ and ρ_+ penalise near and far misses equally

Segmentation error metrics P_k and WD

- P_k [Beeferman et al., 1999]:



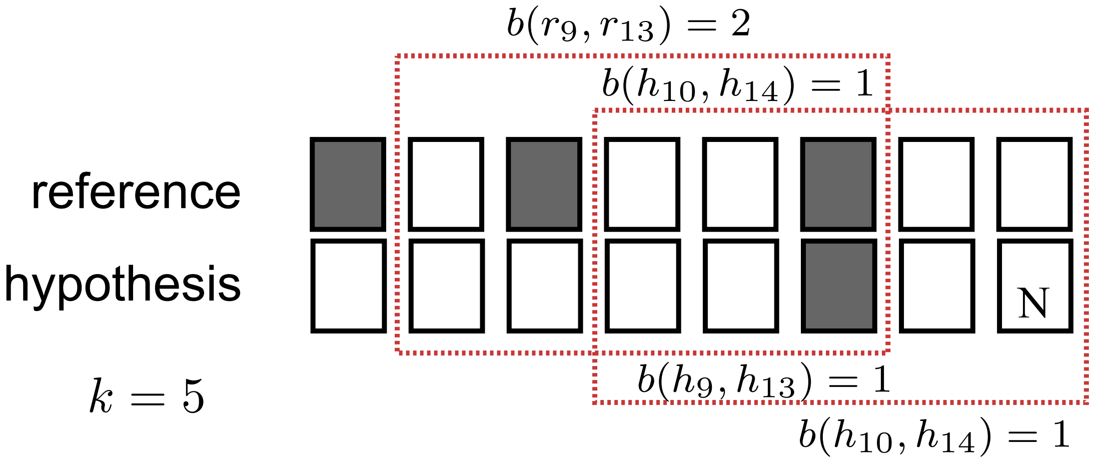
$$P_k(r, h) = \sum_{1 \leq i \leq j \leq N} D_k(i, j)[1 - \delta(a(r_i, r_j), a(h_i, h_j))] \quad (2)$$

- $a(i, j)$ returns 1 if i and j belong to the same PCD
- D_k: distance probability distribution (k assumed to be half the average segment size)

Doug Beeferman, Adam Berger, and John Lafferty. Statistical models for text segmentation. *Machine Learning*, 34:177–210, Feb 1999. URL `http://dx.doi.org/10.1023/A:1007506220214`. 10.1023/A:1007506220214

Segmentation error metrics (ctd)

- *WD* [Pevzner and Hearst, 2002]:



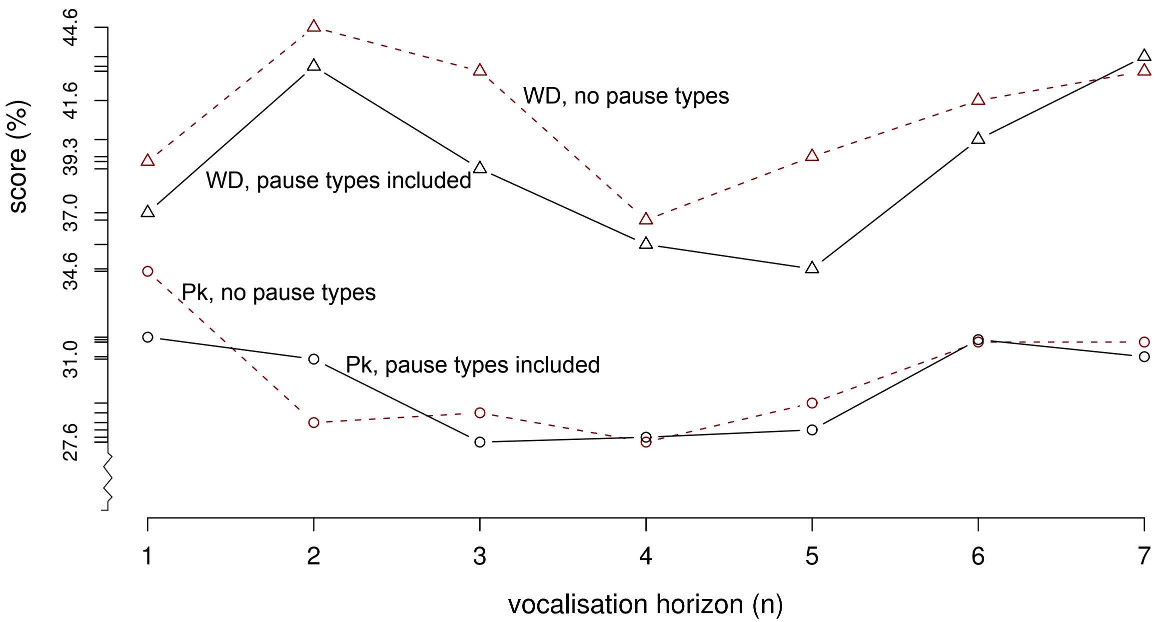
$$\mathrm{WD}(r, h) = \frac{\sum_{i=1}^{N-k}[1 - \delta(b(r_i, r_{i+k}), b(h_i, h_{i+k}))]}{N - k} \tag{3}$$

- $b(i, j)$: the number of PCD boundaries between segments i and j.

Lev Pevzner and Marti A. Hearst. A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28:19–36, Mar 2002. doi: 10.1162/089120102317341756

Some results

All improve considerably on boundary placement baseline (WD$=$ 50.1%, $P_k = 46\%$) by Monte Carlo simulation [Hsueh et al., 2006]



	Pause types included		No pause types	
n	P_k	WD	P_k	WD
1	31.9%	37.0%	34.6%	39.1%
2	31.0%	43.0%	28.4%	44.6%
3	**27.6%**	38.8%	28.8%	42.8%
4	27.8%	35.7%	27.6%	36.7%
5	28.1%	**34.7%**	29.2%	39.3%
6	31.8%	40.0%	31.7%	41.6%
7	31.1%	43.4%	31.7%	42.8%

Some implications

- First-order Markov chains may not be very good models for group conversation;
  - As a further experiment, created a HMM in which $\{b, \neg b\}$ corresponded to hidden states, speaker roles corresponded to observations, and transition and emission probabilities were estimated from the vocalisation matrix.
  - best path hypothesis (Viterbi path) under-predicted, yielding $P_k = 38.2\%$ and $\mathrm{WD} = 41\%$
- Contradicts early opinions on the predictive value of turn taking patterns:

  "turn pattern boundaries [do not appear to be] directly related to discussion topics" [Renals and Ellis, 2003]

The effect of diarisation errors

Usher Institute
Population Health
Sciences & Informatics
www.ed.ac.uk/usher

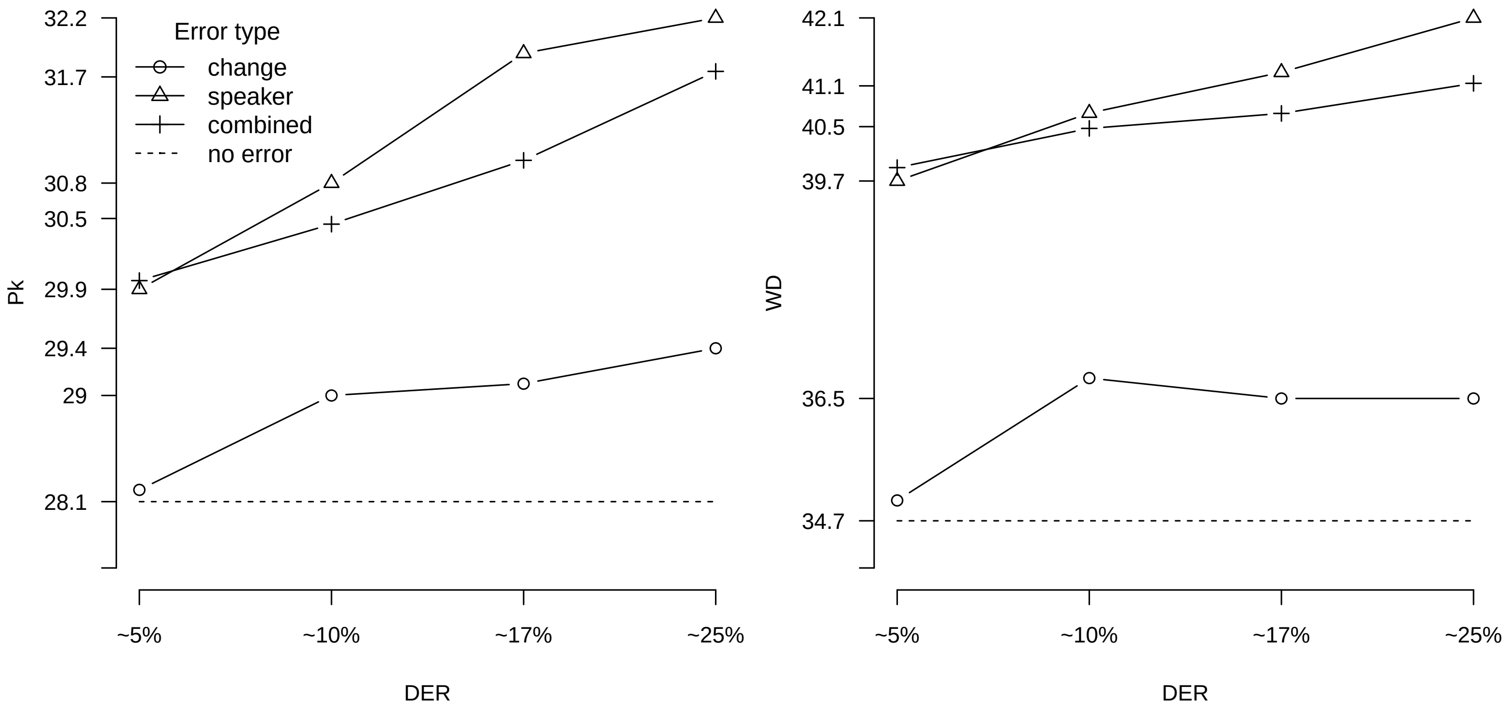


Figure: Segmentation results in P_k (left) and WD (right) for data sets containing diarisation errors. The traced line shows error scores obtained for representation horizon $n = 5$ built on gold standard data (i.e. data containing no diarisation errors).

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Using other ML methods

Table: Performance of segmentation based on different classifiers. Data representation set to a context of 5 vocalisations, including pause type discrimination. Mean number of boundaries per segment fold is 10.

classifier	P_k	WD	# boundaries
NB	**28.1%**	**34.7%**	11.2
k-NN	39.4%	46.3%	7.6
C4.5	41.0%	46.2%	8.0
SVM	34.0%	39.0%	5.2
HMM	38.7%	47.3%	6.0

Comparison with other approaches

Table: Comparison with other meeting segmentation methods. Only the best reported results are presented.

Method	Corpus	Segm. level	P_k	*WD*
LCSeg [Galley et al., 2003]	ICSI	top-level	31.91%	35.88%
LCSeg [Sherman and Liu, 2008]	ICSI	sub-topic	35.29%	42.00%
HMM [Sherman and Liu, 2008]	ICSI	sub-topic	32.70%	39.80%
MAXENT [Hsueh and Moore, 2007]	AMI	top-level	30.00%	**33.00%**
LCSeg [Hsueh and Moore, 2007]	AMI	sub-topic	40.00%	47.00%
MAXENT [Hsueh and Moore, 2007]	AMI	sub-topic	34.00%	36.00%
V_a+MAP+A_f	MDTM	PCD	**27.6%**	34.7%
V_a+MAP+A_f	AMI	sub-topic	**27.67%**	36.00%

Saturnino Luz and Jing Su. The relevance of timing, pauses and overlaps in dialogues: Detecting topic changes in scenario based meetings. In *Proceedings of INTERSPEECH 2010*, pages 1369–1372, Chiba, Japan, 2010. ISCA

Potential Practical applications

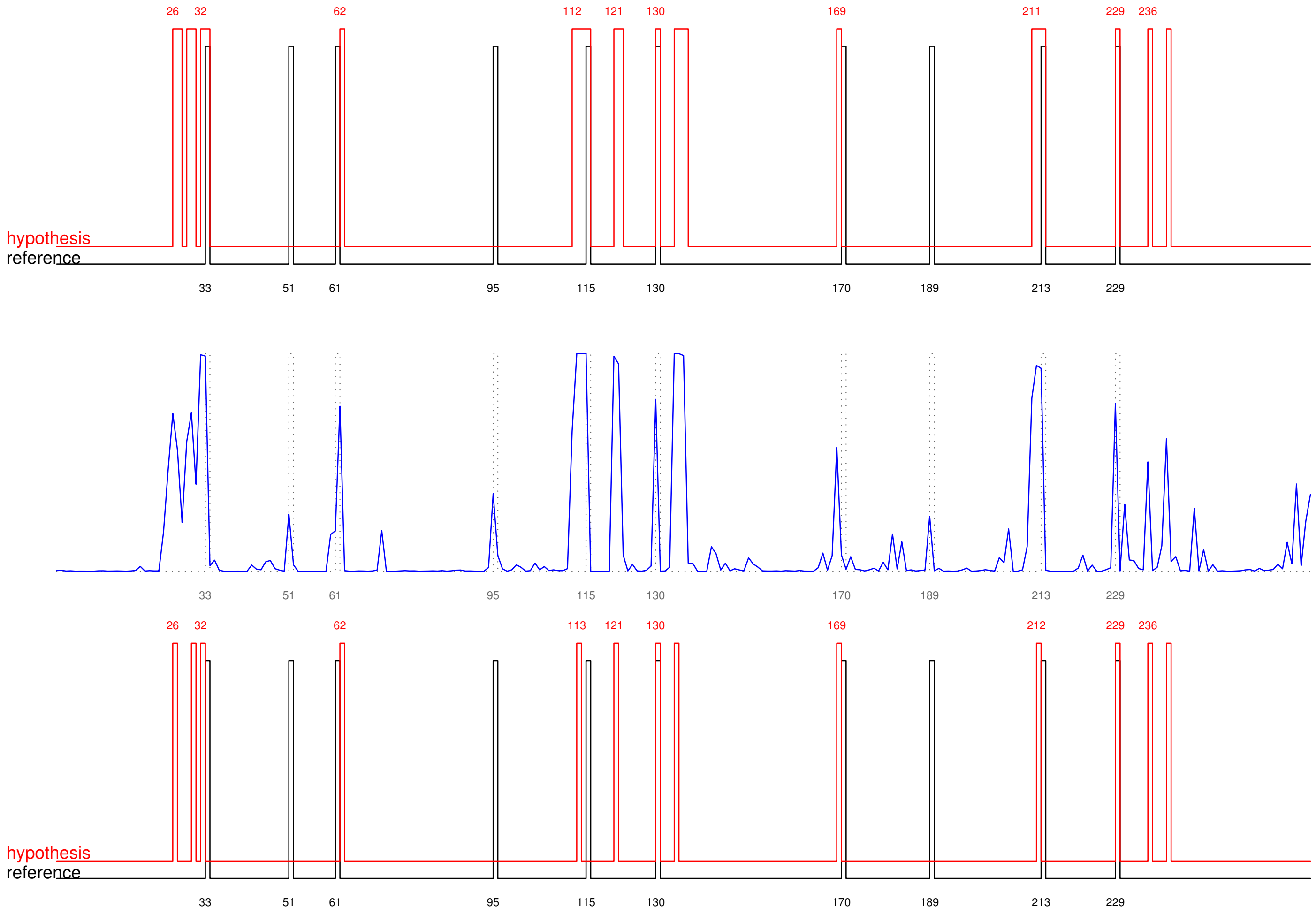
- Storage of PCD discussion records as database objects?

- Certain features of the method presented favour a browsing mode of interaction
  - The method presented above, even though it clearly over-predicts, could usefully support this interaction mode.
- Different categorisation thresholding strategies could be employed to optimise *utility* of the segmentation
- User studies could help determine these parameters

Potential Practical applications

- Storage of PCD discussion records as database objects? NO
  - Current segmentation methods aren't accurate enough.
- Certain features of the method presented favour a browsing mode of interaction
  - The method presented above, even though it clearly over-predicts, could usefully support this interaction mode.
- Different categorisation thresholding strategies could be employed to optimise *utility* of the segmentation
- User studies could help determine these parameters

Segmentation, filtering and categorisation thresholds



Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Lessons

- MDTM segmentation can be achieved with:
  - a simple data representation technique, with roots in CSCW and social psychology research
    +
  - a simple classifier
- Distinguishing among pause types appears to help segmentation
- Extending the vocalisation horizon helps (up to a point)
  - (Naïve Bayes ignores the order of vocalisation events)
- Current segmentation is best suits browsing tasks.

Limitations

- Optimistic assumptions about difficult problems [Jing et al., 2010]:
  - Speech segmentation in noisy environments,
  - diarisation;
- Data collection is difficult:
  - legal issues amd privacy concerns
  - technological and organisational constraints
- Segmentation method tested on “non-standard” corpus; hard to compare our results with other results in the meeting segmentation literature

Speech patterns in Alzheimer's disease

- From characterisation to prediction...

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Yesterday's News?

DAILY NEWSPAPER OF THE YEAR

THE TIMES

SCOTLAND

Top 10 fashion trends

Anna Murphy reports from London's catwalks

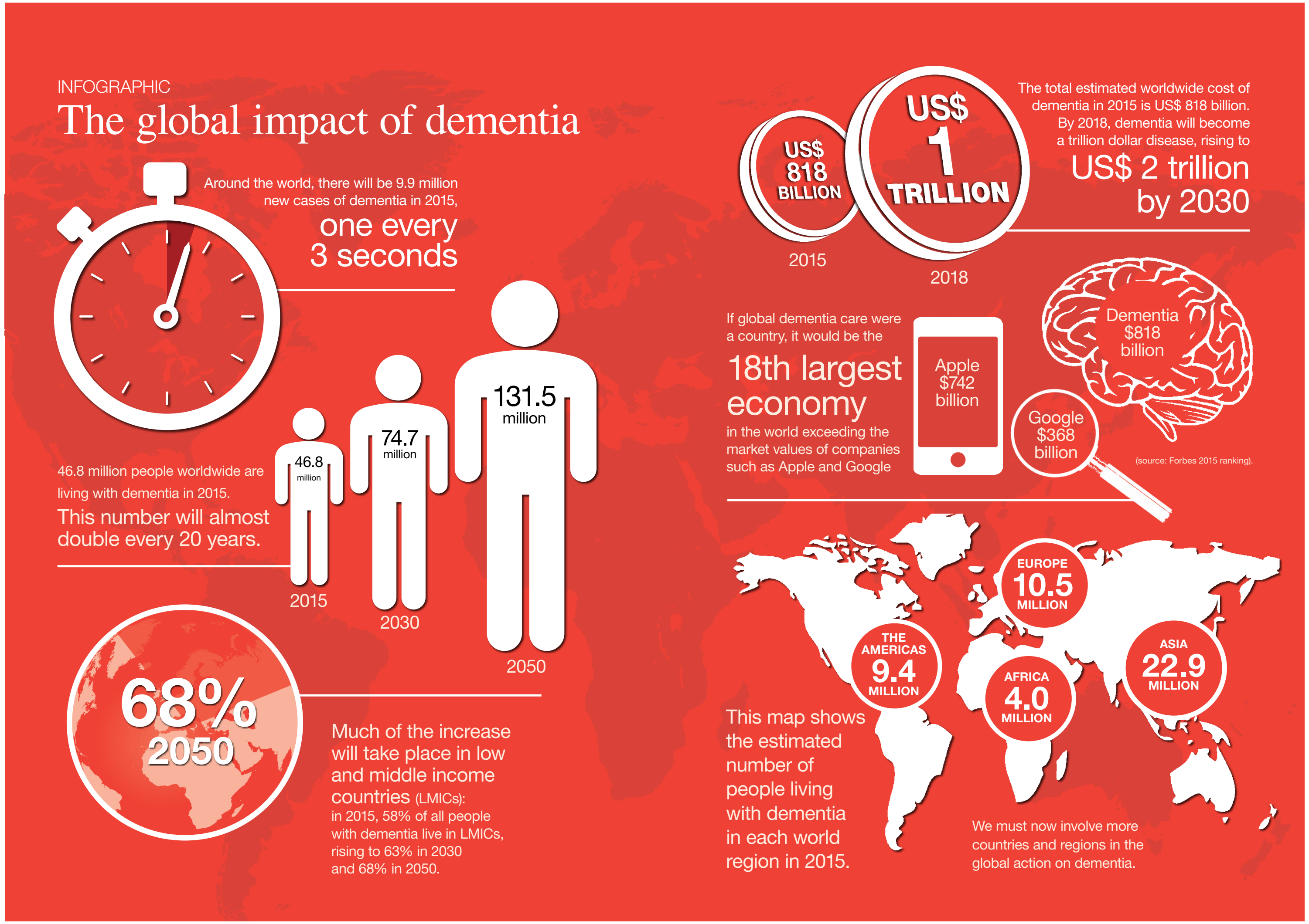
INSIDE TIMES2

Plus Are you wearing commuter trainers?

AI can identify Alzheimer's a decade before symptoms appear

Should drugs

Usher Institute
Population Health
Sciences & Informatics
www.ed.ac.uk/usher



THE UNIVERSITY of EDINBURGH

Martin Prince, Anders Wimo Maëlenn Guerchet, et al. *World Alzheimer Report 2015: The Global Impact of Dementia — Alzheimer's Disease International.* Alzheimers Disease International (ADI), London, 2015

Technology and Communication at the Centre for Dementia Prevention

Characterising Alzheimer's Type Dementia

- Characterisation ("natural history")
  - Neurologic symptoms (gait impairment, buccolingual and limb apraxia, unawareness of memory deficit, agraphesthesia, impaired olfaction, palmomental and snout reflex, etc)
  - Psychiatric symptoms and syndromes (irritability, appetite decrease, low self-esteem, depression, paranoid ideation, general anxiety)
  - Cognitive (linguistic, paralinguistic, memory) changes, performance in neuropsychologic tests, cognitive scores
  - Neuroimaging (PET, fMRI, MRI)
  - Autopsy

Potential Biases

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

TABLE 2. Sources of Potential Bias for Clinicopathologic Correlation Studies of Alzheimer Disease

Patient Characteristics	Clinical Workup	Study Design	Disease Heterogeneity	Pathologic Workup
>Baseline "cognitive reserve" and education-linked factors	>Quantification of "cognition": nonparametric cognitive changes	>Recruitment, inclusion, and exclusion criteria	>Different genetic risk factors at play	>Evaluation and quantification of other pathologies
>Varied access to high-quality health care (diagnostics and therapeutics)	>Quantification of non-AD changes such as cerebrovascular disease	>Cross-sectional vs longitudinal assessments	>Some "atypical" forms of disease	>Focus on complete brain or mainly hippocampus
>Non-AD structural brain comorbidities (cerebrovascular, neurotrauma, etc)	>Cognitive assessment instruments used	>Focus on rare cases or attempting to understand "epidemiological" perspective	>Unknown effects of environmental factors	>Multiple methods to detect AβPs and NFTs
>Emotional and mood disorders	>Individual clinician "thresholds"	>Bias in terms of autopsy rates	>Overlap and interplay between different diseases	>Skew toward end-stage disease at autopsy
>Systemic diseases that affect cognition (metabolic, hormonal, neoplastic, etc.)	>Variation among clinician practices	>Age of individuals in cohort at death	>Specificity of clinical, biomarker, and pathologic features	>Individual pathologist "thresholds"?
>Environmental and behavioral (substance abuse)	>Evolution in assessment methodology over time	>Definitions: "case" and "control" and other terms		>Variation among pathologist practices
>Cohort effects	>Use of biomarkers	>Interval between final clinic evaluation and death		>Accentuation nonhallmark lesions (acetylcholine, synapses)
	>Use of semiquantitative or ordinal variables	>Biostatistical methodology		>Quantitative or ordinal variables

AβP, amyloid β–containing plaque; AD, Alzheimer disease; CPC, clinicopathologic correlations; NFT, neurofibrillary tangle.

THE UNIVERSITY *of* EDINBURGH

Peter T. Nelson, Irina Alafuzoff, Eileen H. Bigio, Constantin Bouras, Heiko Braak, Nigel J. Cairns, Rudolph J. Castellani, Barbara J. Crain, Peter Davies, Kelly Del Tredici, Charles Duyckaerts, Matthew P. Frosch, Vahram Haroutunian, Patrick R. Hof, Christine M. Hulette, Bradley T. Hyman, Takeshi Iwatsubo, Kurt A. Jellinger, Gregory A. Jicha, Enikö Kövari, Walter A. Kukull, James B. Leverenz, Seth Love, Ian R. Mackenzie, David M. Mann, Eliezer Masliah, Ann C. McKee, Thomas J. Montine, John C. Morris, Julie A. Schneider, Joshua A. Sonnen, Dietmar R. Thal, John Q. Trojanowski, Juan C. Troncoso, Thomas Wisniewski, Randall L. Woltjer, and Thomas G. Beach. Correlation of Alzheimer Disease Neuropathologic Changes With Cognitive Status: A Review of the Literature. *Journal of Neuropathology & Experimental Neurology*, 71(5):362–381, May 2012.

ISSN 0022-3069.
doi: 10.1097/NEN.0b013e31825018f7

Diagnosis, early detection

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

- Need for tests that can detect Mild Cognitive Impairment (MCI) and Alzheimer's type Dementia (ATD)
  - characterise impairment
  - monitor interventions/therapy
- Need for methods (other than, but correlated to, neuroimaging) to monitor cognitive status and detect ATD early on:
  - pre-clinical stage, for secondary prevention, or
  - earlier still, for research)
- If such methods exist, they should be catching subtle cognitive changes early on

M. Mortamais, J. A. Ash, J. Harrison, J. Kaye, J. Kramer, C. Randolph, C. Pose, B. Albala, M. Ropacki, C. W. Ritchie, and K. Ritchie. Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, 13(4):468–492, 2017. ISSN 1552-5260. doi: 10.1016/j.jalz.2016.06.2365

Focus on speech and language

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

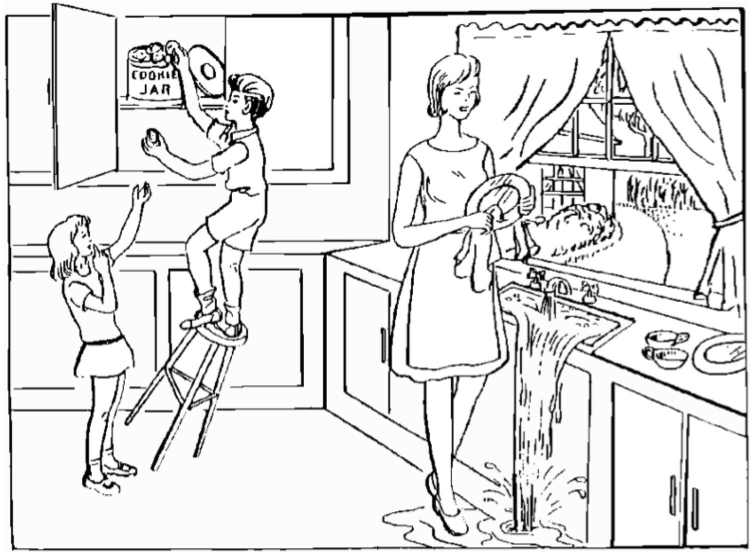
- Much information on cognitive status can be gathered through speech
- Data sources:
  - word tests,
  - narration (scene descriptions),
  - interviews,
  - spontaneous conversations, ...
- Existing Datasets:
  - DementiaBank, the Pitt Dataset,
  - Carolina conversations corpus,
  - ....

Visar Berisha, Shuai Wang, Amy LaCross, and Julie Liss. Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush. *Journal of Alzheimer's Disease*, 45(3):959–963, January 2015. ISSN 1387-2877. doi: 10.3233/JAD-142763

The Pitt Dataset from DementiaBank

Recorded speech data for a number of neuropsychological tests:

- Fluency
- Word recall
- Sentence production
- Cambridge Cookie Theft test:
  - Probable AD speech
  - Normal control speech



Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

JT Becker, F Boiler, OL Lopez, J Saxton, and KL McGonigle. The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594, 1994. doi: 10.1001/archneur.1994.00540180063015

Automatic categorisation of AD and Control speech

- Data, a subset of the Pitt dataset (cookie test):

Control	MCI	Memory	PossibleAD	ProbableAD	Vascular
242	43	3	21	236	5

- Matched by ages and gender:

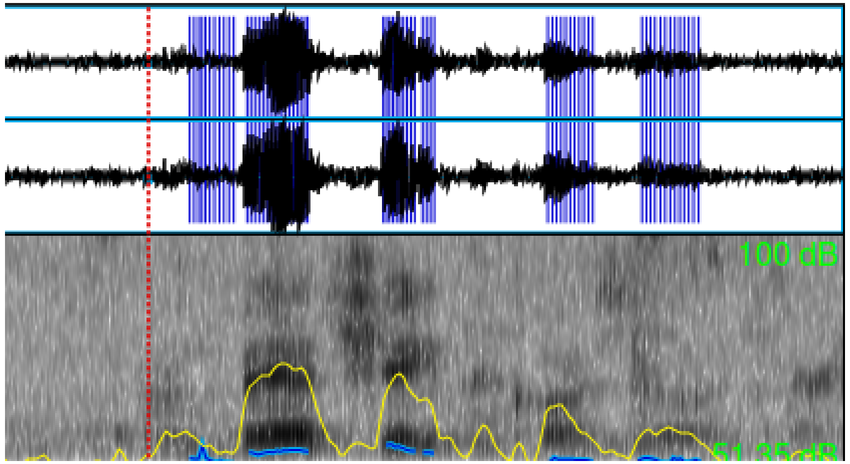
Age interval	Control	AD
$(45, 50]$	0	0
$(50, 55]$	6	6
$(55, 60]$	16	16
$(60, 65]$	35	35
$(65, 70]$	40	40
$(70, 75]$	29	29
$(75, 80]$	10	10

- Extract only prosodic features of speech. Why?

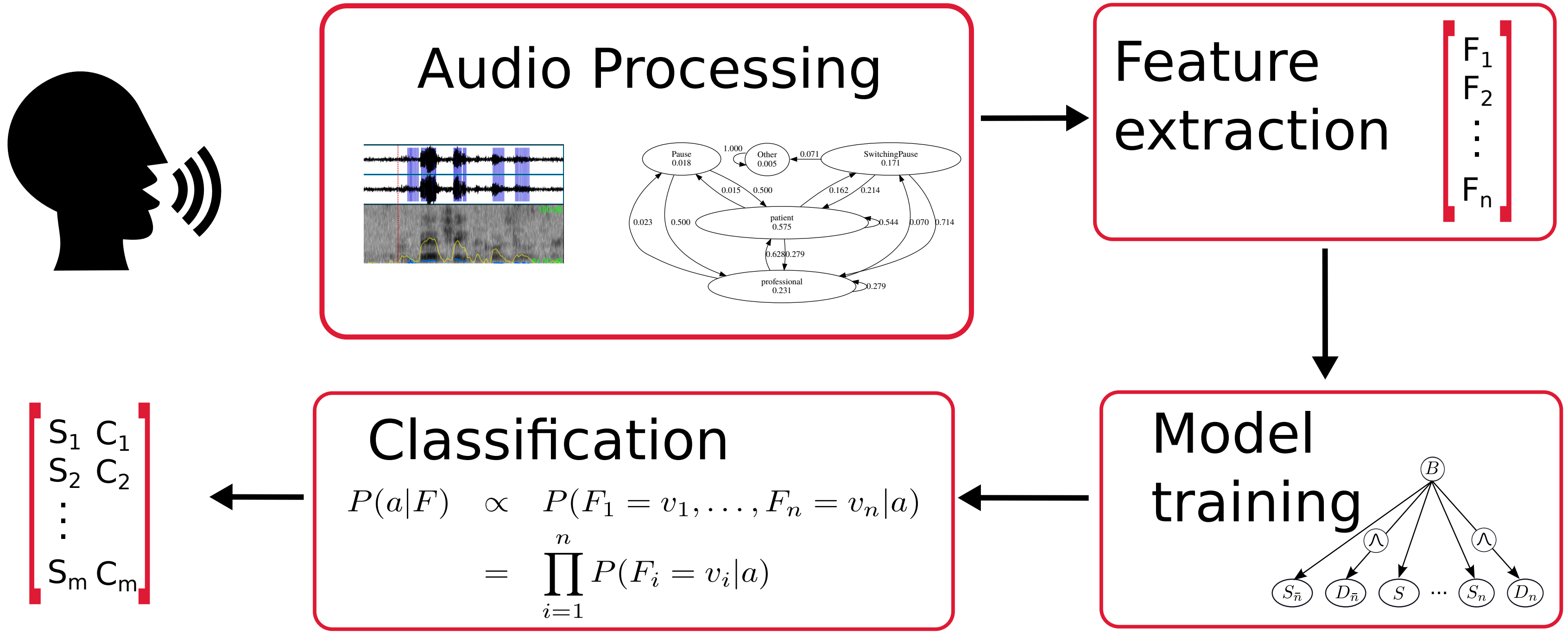
Speech or Language?

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

- There is well documented evidence of linguistic ability impairment (aphasia, and impairment on lexical, syntactic and semantic levels),
- But analysis on these levels requires transcription, etc...
- and we are looking for a method that can be fully automated.
- So we start by investigating prosodic and paralinguistic features of the speech signal:
  - duration of vocalisations and pauses,
  - voice quality measures (e.g. harmonic-to-noise ratio, pitch, loudness peaks, speaking rate, etc)

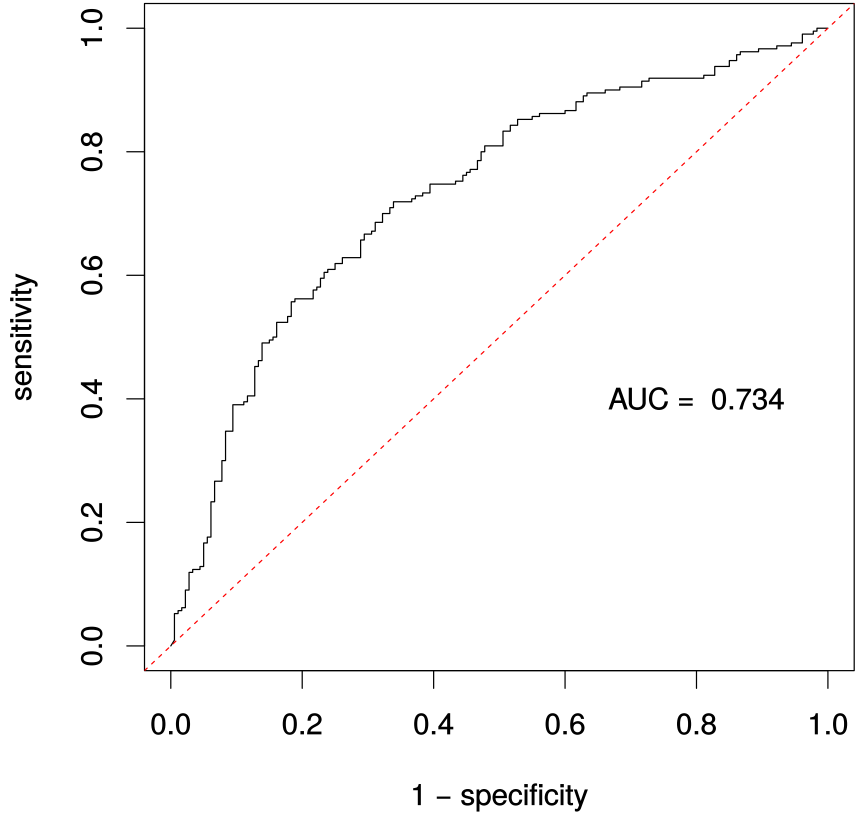


Analysis pipeline



Results (Pitt dataset)

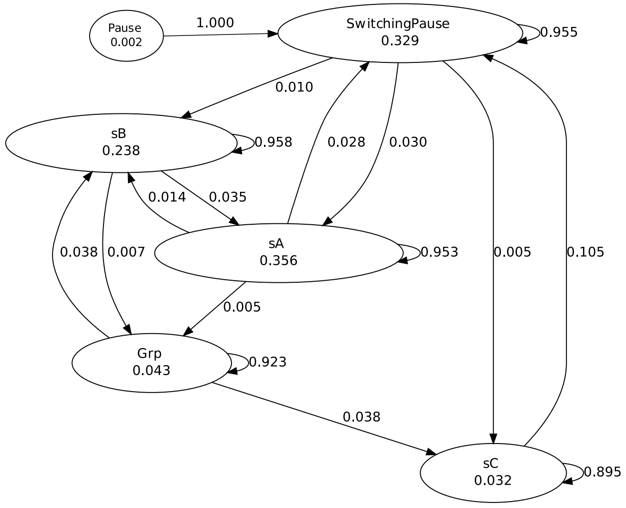
- Overall accuracy of 68%,
- F_1 scores of 70% for the control class and 64% for the ATD class
  - ("gold standard" = diagnostic established through a combination of neuropsychological and neurologic tests)

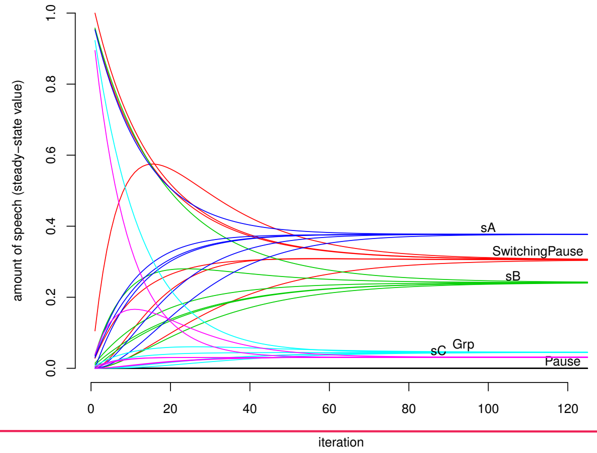


Saturnino Luz. Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data.
In *Computer Based Medical Systems*, pages 45–46. IEEE Press, 2017

From individual speech to interaction

- Most studies in this area have focused on individual speech patterns of the patient.
- How about analysing patterns in dialogue?
- Background:
  - Inspired by Jaffe's and Feldstein's pioneering work on conversation timing
  - Successful applications to group performance prediction
  - and meeting segmentation





THE UNIVERSITY of EDINBURGH

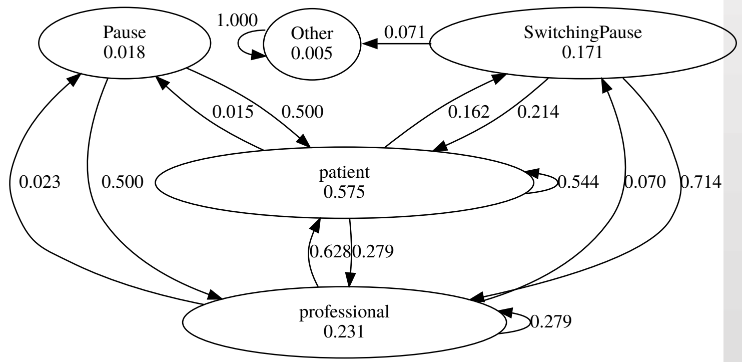
Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

Joseph Jaffe and Stanley Feldstein. *Rhythms of dialogue.* Personality and Psychopathology. Academic Press, New York, 1970

Saturnino Luz. Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI'13, pages 575–582. ACM, 2013

Patient interviews

- Spontaneous, conversational speech from the Carolina Conversations Collection (CCC)
- Hypothesis: AD patient dialogues show identifiable patterns during dialogue interactions, such as disrupted turn taking and differences in speech rate.
- Subset of CCC:
  - 21 patients with a diagnosis of Alzheimer's disease (15 females, 6 males),
  - and 17 patients (12 females, 5 males) with other diseases (excluding neuropsychological conditions)
  - matching age range and gender frequencies.
- Representation:
  - vocalisation graphs (VGO)
  - vocalisation graphs + speech rate (VGS)



Machine Learning

- A simple boosting model was trainied using decision stumps as weak learners.
- The training process consisted of 10 iterations whereby, for each training instance (x_i), a weak classifier $\hat{f}_m$ was fitted using weights on the data
- final classifiation decision (additive logistic regression) following the Real Adaboost algorithm:

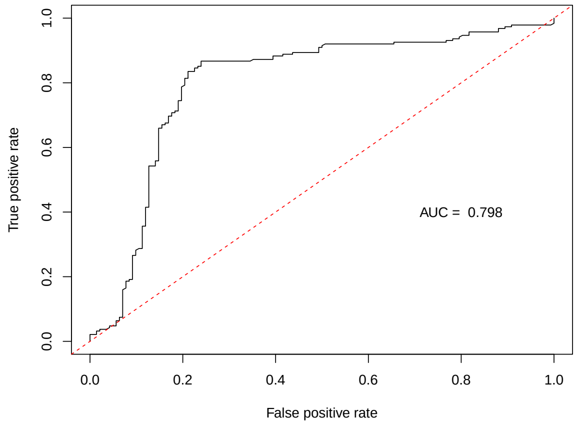
$$\hat{F}(x) = sign\left[\sum_{m=1}^{M} \hat{f}_m(x)\right] \quad (4)$$

Results for dialogues

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

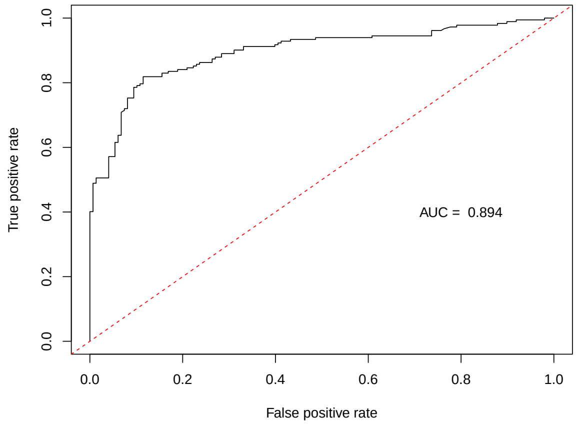
AD detection results for the VGO data representation scheme.

AD		non-AD	
Accuracy_μ	0.812	Accuracy_μ	0.714
Precision_μ	0.765	Precision_μ	0.769
Recall_μ	0.812	Recall_μ	0.714
$F_{1,\mu}$	0.788	$F_{1,\mu}$	0.741
Precision_M	0.667	Precision_M	0.792
Recall_M	0.722	Recall_M	0.729
$F_{1,M}$	0.685	$F_{1,M}$	0.721
Overall accuracy (LOOCV): 0.811			



Results for the VGS data representation scheme.

AD		non-AD	
Accuracy_μ	0.882	Accuracy_μ	0.769
Precision_μ	0.833	Precision_μ	0.833
Recall_μ	0.882	Recall_μ	0.769
$F_{1,\mu}$	0.857	$F_{1,\mu}$	0.800
Precision_M	0.796	Precision_M	0.708
Recall_M	0.833	Recall_M	0.708
$F_{1,M}$	0.811	$F_{1,M}$	0.700
Overall accuracy (LOOCV): 0.865			



Some comparisons

Compared accuracy results obtained with different classification algorithms, on VGS-based datasets.

Classification method	Accuracy (LOOCV)
Logistic regression	75.7%
Real Adaboost	86.5%
Decision trees	86.5%
SVM	83.7%
Random forests	81.1%

Future

- Monitor health by analysing everyday (inter)activity data
- Mobile and wearable devices, a multitude of sensors, companion robots, ...
- Current focus: monitoring cognitive function (applications in elderly care)

SAM Supporting Active Ageing through multimodal coaching

- Incorporation of speech and interaction data into research on early cognitive signs of dementia.

Challenges

- Scarcity of spontaneous speech/interaction data, particularly longitudinal data
- Challenges in data pre-processing
  - segmentation,
  - diarisation,
  - feature extraction.
- Lack of theoretical foundations for “interaction analytics”.
- But elements of such foundations may exist:
  - Theories of alignment/entrainment phenomena in communiation (from psycholinguistics)
  - Interaction turns may be modulations on CPG patterns
    (building on an idea by Björn Lindblom and advances in neuroscience)

Ongoing work on AD prediction

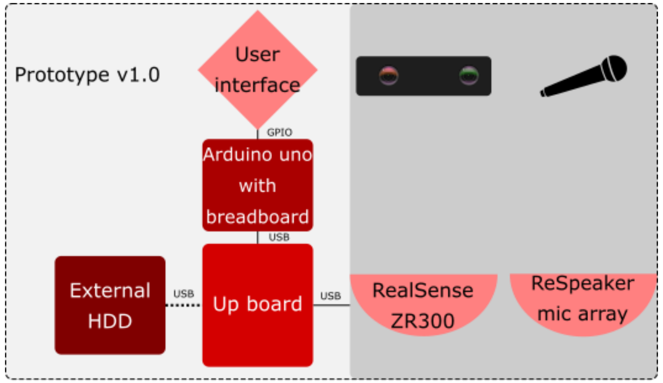
- Started collecting dialogue samples in a visuospatial task from the PREVENT cohort
- PREVENT explores differences in the brain and cognitive function in healthy people in mid-life (aged 40-59). People are grouped into high, mid and low risk based on their family history and APOE status
- Goal: to detect subtle signs of AD in mid-life through analysis of speech

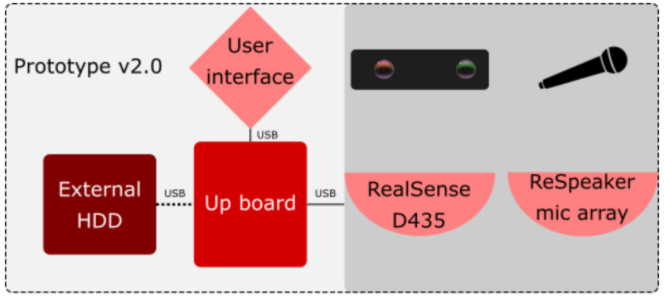
Craig W Ritchie and Karen Ritchie. The PREVENT study: a prospective cohort study to identify mid-life biomarkers of late-onset alzheimer's disease. *BMJ open*, 2(6):e001893, 2012

Ongoing work on work communication in healthcare

Usher Institute
Population Health Sciences & Informatics
www.ed.ac.uk/usher

- INCA and VICO projects: Analysis of Patient-GP communication
- Bespoke hardware development:
  - The CUstom Secure ReCOrder (CUSCO)
  - The device supports the collection of dialogues by recording audio and video streams with a quality suitable for speech processing and NLP
  - Data encrypted on collection.
  - Developed with a modular software architecture for easier adaptation to different scenarios and sensors.
- Data collection in progress





References

Robert F. Bales. *Interaction Process Analysis: A Method for the Study of Small Groups.* Addison-Wesley, Cambridge, Mass., 1950.

JT Becker, F Boiler, OL Lopez, J Saxton, and KL McGonigle. The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594, 1994. doi: 10.1001/archneur.1994.00540180063015.

Doug Beeferman, Adam Berger, and John Lafferty. Statistical models for text segmentation. *Machine Learning*, 34:177–210, Feb 1999. URL http://dx.doi.org/10.1023/A:1007506220214. 10.1023/A:1007506220214.

Visar Berisha, Shuai Wang, Amy LaCross, and Julie Liss. Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush. *Journal of Alzheimer's Disease*, 45(3):959–963, January 2015. ISSN 1387-2877. doi: 10.3233/JAD-142763.

P.T. Brady. A statistical analysis of on-off patterns in 16 conversations. *The Bell System Technical Journal*, 47:73–91, 1968.

James M. Jr. Dabbs and Barry Ruback. Dimensions of group process: Amount and structure of vocal interaction. *Advances in Experimental Social Psychology*, 20(123–169), 1987.

Michel Galley, Kathleen R. McKeown, Eric Fosler-Lussier, and Hongyan Jing. Discourse segmentation of multi-party conversation. In Erhard Hinrichs and Dan Roth, editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 562–569, 2003. URL http://www.aclweb.org/anthology/P03-1071.pdf.

Pei-Yun Hsueh and Johanna D. Moore. Combining multiple knowledge sources for dialogue segmentation in multimedia archives. In *Proceedings of the 45th Annual Meeting of the ACL*, pages 1016–1023, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

Pei-yun Hsueh, Johanna D. Moore, and Steve Renals. Automatic segmentation of multiparty dialogue. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 273–277. ACL Press, 2006.

Joseph Jaffe and Stanley Feldstein. *Rhythms of dialogue.* Personality and Psychopathology. Academic Press, New York, 1970.

Su Jing, Bridget Kane, and Saturnino Luz. Automatic meeting participant role detection by dialogue patterns. In *Development of Multimodal Interfaces: Active Listening and Synchrony*, number 5967 in Lecture Notes in Computer Science, pages 314–327. Springer, 2010. doi: 10.1007/978-3-642-12397-9_27.

George H. John and Pat Langley. Estimating continuous distributions in Bayesian classifiers. In Besnard, Philippe and Steve Hanks, editors, *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI'95)*, pages 338–345, San Francisco, CA, USA, August 1995. Morgan Kaufmann Publishers.

Bridget Kane and Saturnino Luz. Multidisciplinary medical team meetings: An analysis of collaborative working with special attention to timing and teleconferencing. *Computer Supported Cooperative Work (CSCW)*, 15(5):501–535, 2006. doi: 10.1007/s10606-006-9035-y.

Bridget Kane and Saturnino Luz. Achieving diagnosis by consensus. *Computer Supported Cooperative Work (CSCW)*, 18(4):357–391, 2009. doi: 10.1007/s10606-009-9094-y.

Bridget Kane and Saturnino Luz. Information sharing at multidisciplinary medical team meetings. *Group Decision and Negotiation*, 20:437–464, 2011. doi: 10.1007/s10726-009-9175-9.

Bridget Kane and Saturnino Luz. "Do no harm": Fortifying MDT collaboration in changing technological times. *International Journal of Medical Informatics*, 82(7):613–625, July 2013. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2013.03.003.

Bridget T. Kane, Pieter J. Toussaint, and Saturnino Luz. Shared decision making needs a communication record. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 79–90, New York, NY, USA, 2013. ACM. doi: 10.1145/2441776.2441788.

Saturnino Luz. The non-verbal structure of patient case discussions in multidisciplinary medical team meetings. *ACM Transactions on Information Systems*, 30(3):17:1–17:24, 2012.

Saturnino Luz. Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI'13, pages 575–582. ACM, 2013.

Saturnino Luz. Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data. In *Computer Based Medical Systems*, pages 45–46. IEEE Press, 2017.

Saturnino Luz and Bridget Kane. Perspectives on intelligent systems support for multidisciplinary medical teams. In *AAAI Spring Symposium Series*, pages 272–275, 2016. URL http://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12741/11964.

Saturnino Luz and Jing Su. The relevance of timing, pauses and overlaps in dialogues: Detecting topic changes in scenario based meetings. In *Proceedings of INTERSPEECH 2010*, pages 1369–1372, Chiba, Japan, 2010. ISCA.

M. Mortamais, J. A. Ash, J. Harrison, J. Kaye, J. Kramer, C. Randolph, C. Pose, B. Albala, M. Ropacki, C. W. Ritchie, and K. Ritchie. Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, 13(4):468–492, 2017. ISSN 1552-5260. doi: 10.1016/j.jalz.2016.06.2365.

Peter T. Nelson, Irina Alafuzoff, Eileen H. Bigio, Constantin Bouras, Heiko Braak, Nigel J. Cairns, Rudolph J. Castellani, Barbara J. Crain, Peter Davies, Kelly Del Tredici, Charles Duyckaerts, Matthew P. Frosch, Vahram Haroutunian, Patrick R. Hof, Christine M. Hulette, Bradley T. Hyman, Takeshi Iwatsubo, Kurt A. Jellinger, Gregory A. Jicha, Enikö Kövari, Walter A. Kukull, James B. Leverenz, Seth Love, Ian R. Mackenzie, David M. Mann, Eliezer Masliah, Ann C. McKee, Thomas J. Montine, John C. Morris, Julie A. Schneider, Joshua A. Sonnen, Dietmar R. Thal, John Q. Trojanowski, Juan C. Troncoso, Thomas Wisniewski, Randall L. Woltjer, and Thomas G. Beach. Correlation of Alzheimer Disease Neuropathologic Changes With Cognitive Status: A Review of the Literature. *Journal of Neuropathology & Experimental Neurology*, 71(5):362–381, May 2012. ISSN 0022-3069. doi: 10.1097/NEN.0b013e31825018f7.

Lev Pevzner and Marti A. Hearst. A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28:19–36, Mar 2002. doi: 10.1162/089120102317341756.

Martin Prince, Anders Wimo Maëlenn Guerchet, et al. *World Alzheimer Report 2015: The Global Impact of Dementia — Alzheimer's Disease International.* Alzheimers Disease International (ADI), London, 2015.

S. Renals and D. Ellis. Audio information access from meeting rooms. In *Procs. of the Internacional Conference on Acoustics, Speech, and Signal Processing*, volume IV, pages 744–747. IEEE, 2003.

Craig W Ritchie and Karen Ritchie. The PREVENT study: a prospective cohort study to identify mid-life biomarkers of late-onset alzheimer's disease. *BMJ open*, 2(6):e001893, 2012.

Abigail J. Sellen. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*, 10(4):401–444, 1995. doi: 10.1207/s15327051hci1004_2. URL http://www.leaonline.com/doi/abs/10.1207/s15327051hci1004_2.

M. Sherman and Y. Liu. Using hidden Markov models for topic segmentation of meeting transcripts. In *Proceedings of the IEEE Spoken Language Technology Workshop*, pages 185–188, 2008.