

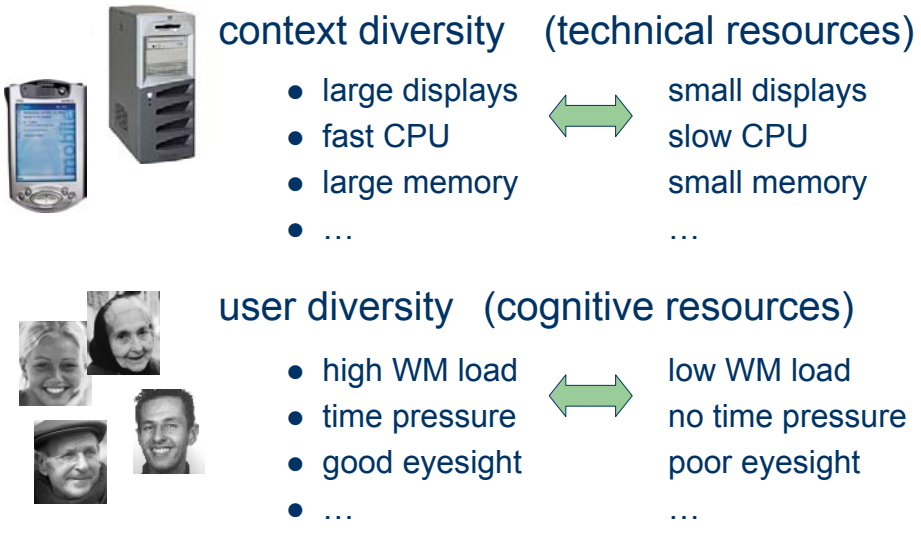
Implicit Feedback for User-Adaptive Systems by Analyzing the Users' Speech

Frank Wittig / Christian Müller
Saarland University,
Germany
{wittig, cmueller}@cs.uni-sb.de

Outline

- Motivation: Universal Usability
- M3I
- Speech as a Source for UM
- Two-Layered ML Recognition Approach
- Summary / Current and Future Work

Universal Usability



M3I

- a **M**obile **M**ulti-modal and **M**odular Interface
- developing a framework for resource-adaptive mobile multi-modal systems
 - recognizing both technical and cognitive resource limitations
 - adapting the systems behaviour accordingly
- applications
 - mobile pedestrian navigation system
 - shopping assistant
 - museum guide

This talk

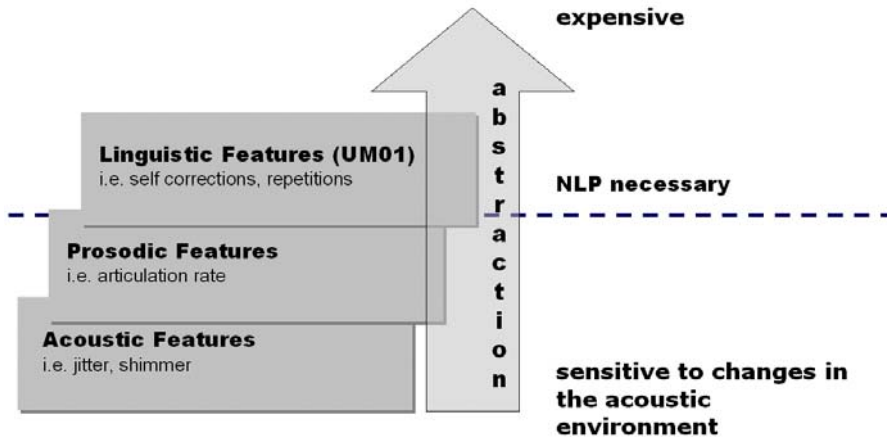
- a **Mobile Multi-modal and Modular Interface**
- developing a framework for resource-adaptive mobile multi-modal interfaces
 - recognizing both technical and cognitive resource limitations
 - adapting the systems behaviour accordingly
- applications
 - mobile pedestrian navigation system
 - shopping assistant
 - museum guide

Simplification

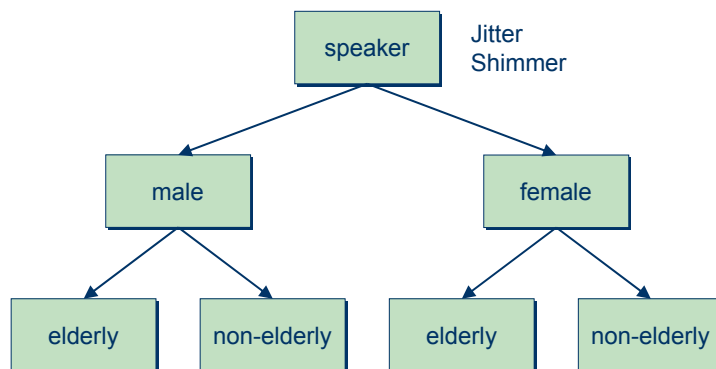
- The notion of cognitive resources is restricted to the distinction between

„average aged adults“  the elderly

Speech as a Source for UM



Classification Tree

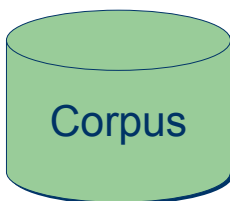


„voices of men and women age differently“

Two-Layered Machine Learning Approach

- Classification Layer
 - artificial neural network
 - classifies a given speaker (uncertain)
- Meta-Reasoning Layer
 - dynamic Bayesian network
 - treats uncertainty
 - models gender specific aging
 - improves model incrementally

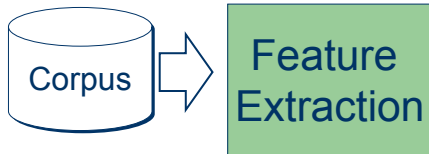
Classification Layer



Number of speakers

male	586
female	442
elderly	676
non-elderly	352

Classification Layer



- Jitter
 - Shimmer
- (11 features in total)

Classification Layer



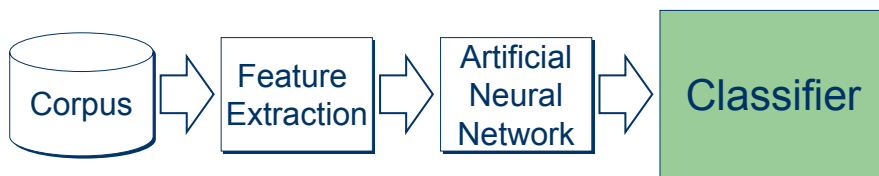
- performed best in initial study compared with C 4.5 decision tree induction, k-nearest neighbors, naive Bayes, and support vector machines

Alternative Classifiers

Prediction Accuracy (10-fold CV, WEKA):

	C4.5	ANN	kNN	NB	SVM	BL
Gender	88.07	93.14	91.55	81.64	86.07	57.84
Age	85.30	92.88	88.10	70.27	79.26	64.16

Classification Layer



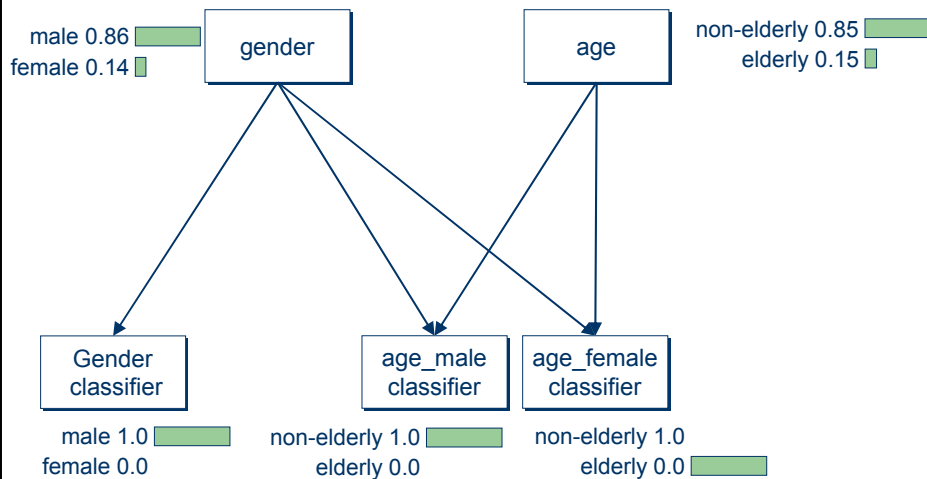
Classification Layer

Classifier

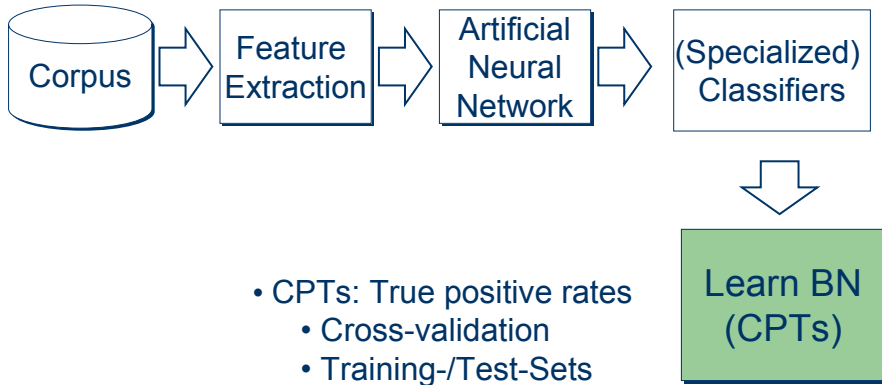
Classification accuracy
(10-fold cross validation)

male	~96%
female	~89%
elderly	~88%
non-elderly	~96%

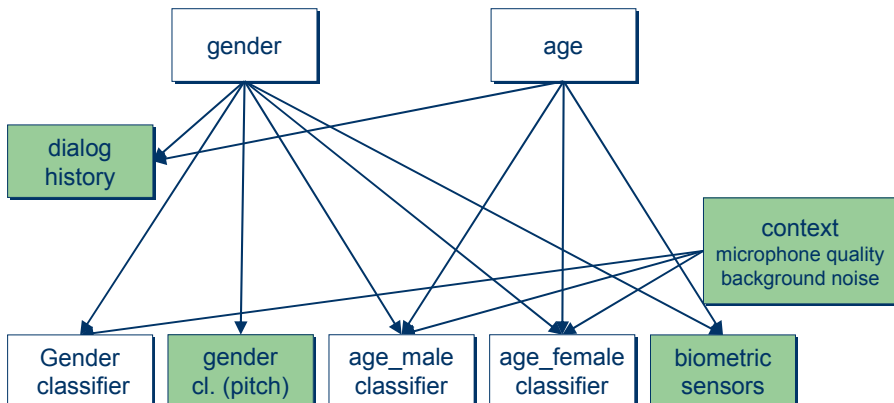
Meta-Reasoning Layer



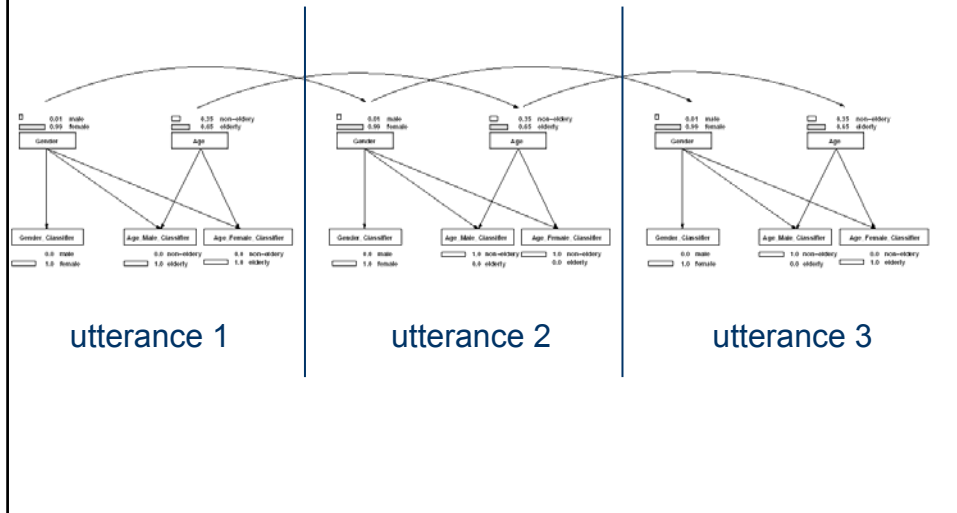
Meta-Reasoning Layer



Meta-Reasoning Layer (future extensions)



Continuously Updating the Model



Summary and Future Work

- two-layered machine-learning approach for estimating age and gender of a speaker
 - classification layer (ANN):
 - high accuracy using two acoustic measures jitter and shimmer.
 - meta-reasoning layer (DBN):
 - modeling uncertainty and gender-specific aging
 - incrementally updating the user model
- the approach is implemented in a client/server architecture
- main issues of (near) future work
 - integration into testbed applications
 - pedestrian navigation and shopping/museum guide
 - developing adaptation strategies

Thank you very much
for your attention !