Speech Recognition of Aged Voice in the AAL Context: Detection of Distress Sentences

F. Aman, **M. Vacher**, S. Rossato and F. Portet

ANR-2010-TECS-012 ANR

Outline

General Facts and Context

Growing number of ageing people

- rise of life expectancy;
- wishes of the people to stay at home;
- overflow of care institutions.

Health Smart Homes: A solution for elderly people to live independently at home?

- Autonomy:
  $\Longrightarrow$ homecare.
- Home automation for disability compensation:
  $\Longrightarrow$ light control, events reminder,
  $\Longrightarrow$ facilities for contact with the family.
- Security:
  $\Longrightarrow$ detection of distress situations...

Interest of Speech Technologies

- Targeted audience: aged people with loss of autonomy
  - Reduced autonomy
  - Growing social isolation
  - Chronic and degenerative diseases (Alzheimer)
- Problems:
  - Complex interfaces
  - Lack of familiarity of this population with new technologies

Interaction through voice:

- Natural human / machine interaction
- Automatic speech recognition of the elderly voice
  - Call a relative for help
  - Vocal orders for home automation

Challenges

- distant speech recognition conditions
  - reverberation, etc.
- presence of noise
  - blind source separation
- spontaneous speech
- aged voice
  - the language of the elderly becomes a specific language (change of use due to their cognitive evolution)
- imprecise production of consonants, tremors and slower articulation, anatomical evolution, mouth noise, ...
- emotion in the voice in distress situation conditions

Specificity of elderly voice

- VIPPERLA's study for English language:
  - audio recordings of the proceedings of the Supreme court of the USA
  - defence speech of 7 judges
  - recording of the same speaker every year
  - limits of this study: persons familiar to deliver speeches in public
- Conclusion of this study:
  - modification of the fundamental frequency
  - lack of energy
  - instability in the production of certain consonants
  - increasing noise

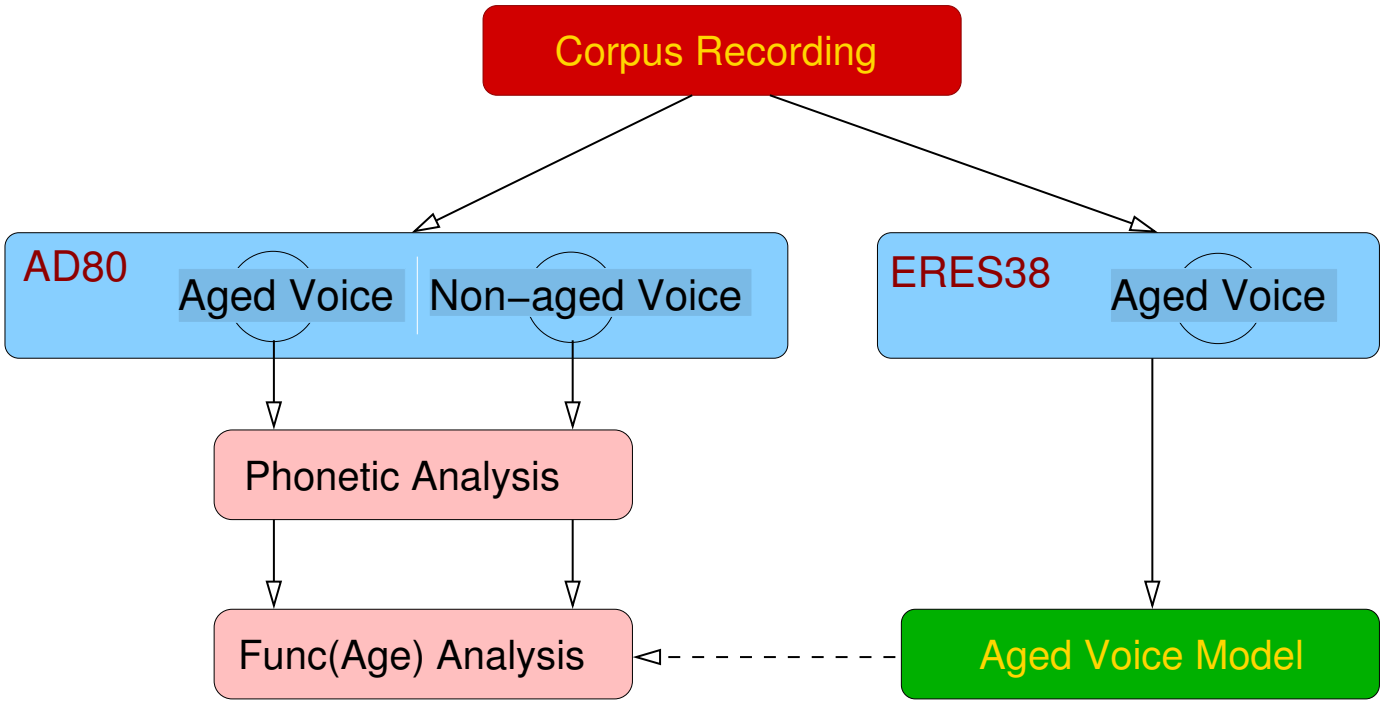
Evaluation of the impact of the aging voice on the ASR performance

- Few existing corpora of elderly voice in French
  - spontaneous speech
  - non adapted to home automation
  - non adapted to distress calls

- Necessity of recording and annotating of new corpora

- Difficulties of elderly recording:
  - at their own home
  - fears with respect to the use of the recorded data
  - tiredness and lack of concentration of elderly people

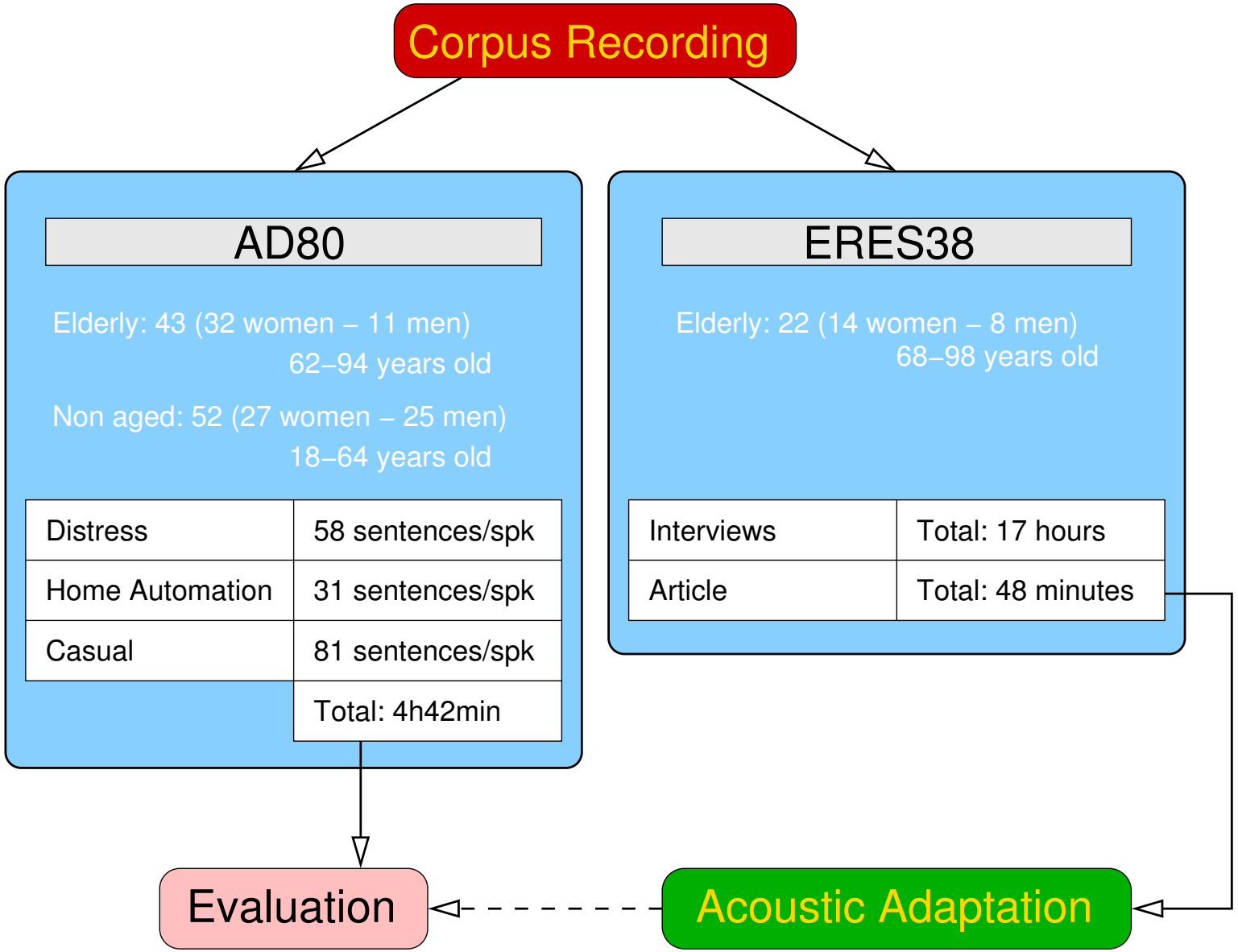
Ageing voice: Methodology



Automatic speech Recognizer:

- Sphinx3 engine
- Generic acoustic models: French corpus BREF120
- Specific adaptation: MLLR (Maximum Likelihood Linear Regression)

Resulting corpora (read sentences)



Example of read sentences (AD80)

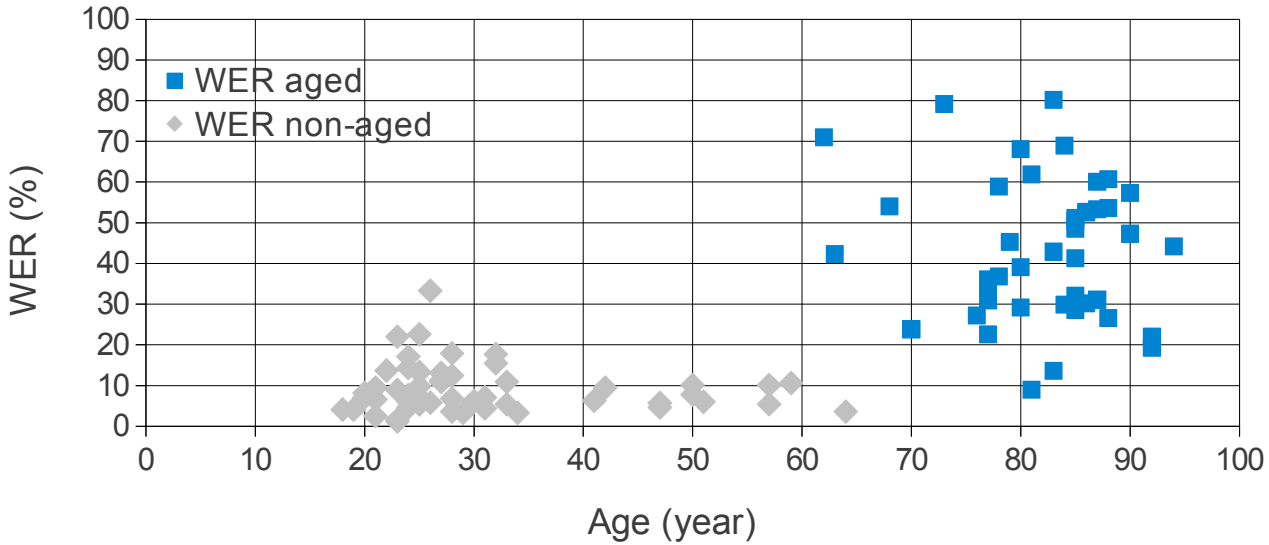
- Home automation
  - Appelle quelqu'un e-lio !
  - e-lio appelle ma fille !
  - e-lio appelle les secours !
  - e-lio tu peux téléphoner au SAMU ?
  - e-lio appelle du secours !
- Casual
  - Bonjour madame !
  - J'allume la lumière !
  - J'ai ouvert la porte !
  - Où sont mes lunettes !
  - Ça va très bien !
- Distress

  (choosen after interviews with elderly and professionnal carers)
  - Qu'est-ce qui m'arrive !
  - Oh là! Je saigne ! Je me suis blessé !
  - Je peux pas me relever !
  - Aidez-moi !
  - Je suis tombé !

ASR toolkit: Sphinx3

- **Generic acoustic model:**
  - Context dependent
  - 3-state HMM
  - 13 MFCC coefficients , Δ, $\Delta\Delta$
  - Training with 100 hours of the BREF120 French corpus
- **Language models:**
  - 3-gram LM
  - Interpolation of:
    - Generic LM (10%): French Gigaword, 1-gram, 11,018 words
    - Specialized LM (90%): distress sentences, 3-gram, 88 words

First Experiment (generic acoustic models)



WER as a function of age for aged and non-aged groups

- Average WER:
  - Non elderly: 7.1% (σ=6%)
  - Elderly: 43.5% (σ=17.3%)

Conclusion:

- Standard acoustic models unadapted to elderly
- High variability in the case of aged people

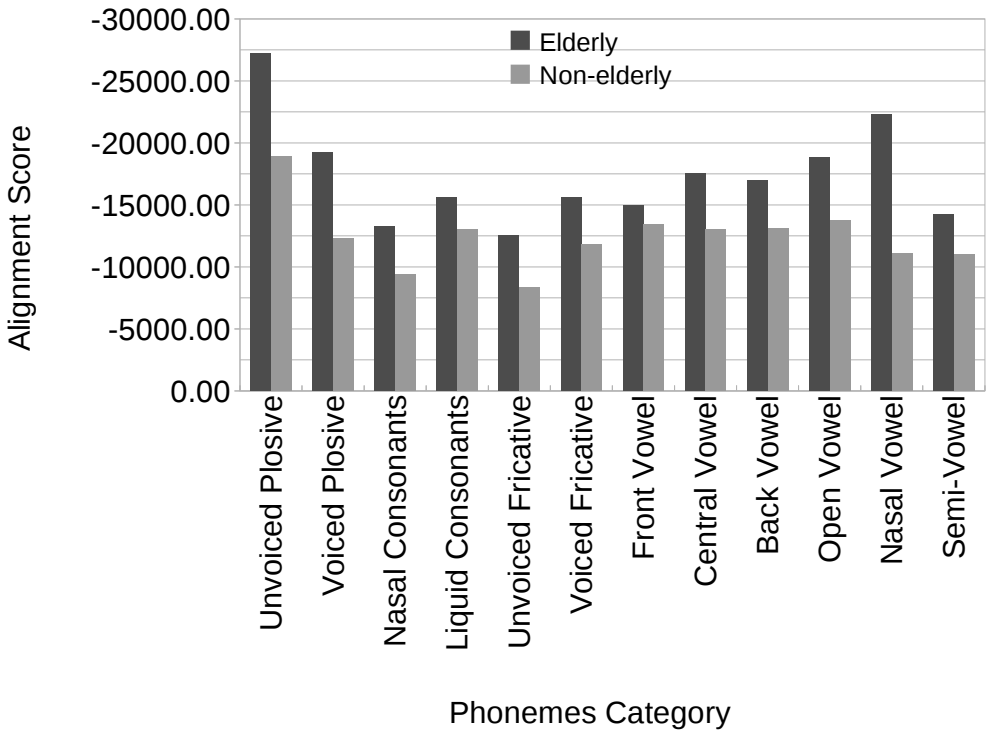
Second Experiment (1/2)

unvoiced plosive	p, t, k
voiced plosive	b, d, g
nasal consonant	m, n, ŋ, ɲ
liquid consonant	l
unvoiced fricative	f, s, ʃ
voiced fricative	v, z, ʒ, ʁ
front vowel	i, e, ɛ
central vowel	y, ø, œ, ə
back vowel	u, o, ɔ
open vowel	a, ɑ
nasal vowel	ẽ, ã, õ, œ̃
semi-vowel	ɥ, j, w

Table: Phoneme categories for French (IPA symbols)

Second Experiment (2/2)

(generic acoustic models)



Forced alignment scores by phoneme categories

Conclusion:

Most affected categories for elderly group (relative differences): nasal vowels (-100%), voiced plosive (-56%), unvoiced fricatives (-50%) and nasal consonants (-41%)

Third Experiment

MLLR adaptation with the speech of 22 elderly speakers of the ERES38 corpus:

- **generic acoustic model for elderly**
- average WER after acoustic adaptation: 14.5%

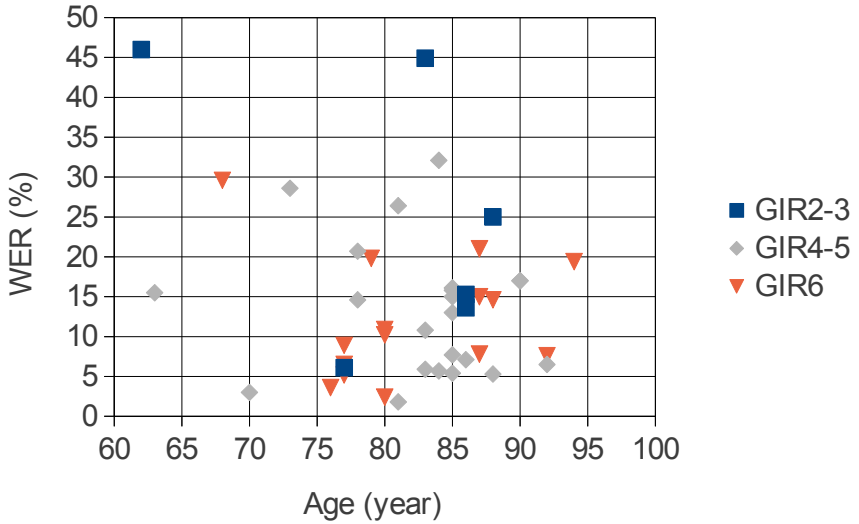
Problem:

- adapted acoustic models reduce the WER significantly for the most important part of the speakers
- some speakers are badly recognized

Influence of the level of dependence

GIR: French scale for dependence characterisation

- GIR6: no dependence
- GIR4-5: small dependence
- GIR2-3: important dependence
- GIR1: full dependence



Conclusion:

- important dispersion for each GIR group
- highest dispersion for most dependent people
- necessity of more precise criterion (related to physical degradation?)

Detection of emergency calls

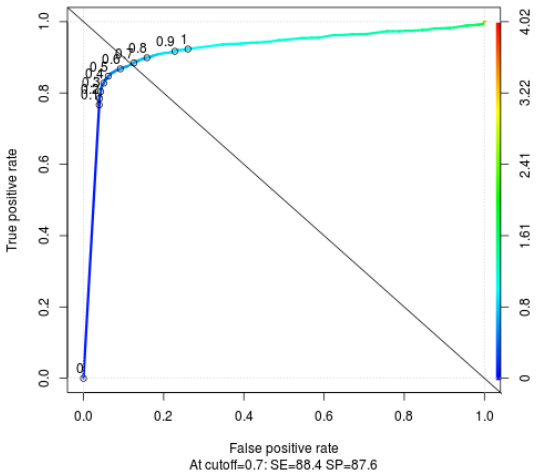
Examples of calls:

Au secours	Appelle quelqu'un e-lio
Qu'est-ce qui m'arrive	e-lio appelle les secours
Aidez moi	e-lio appelle quelqu'un

- Levenshtein distance
- distance between the hypothesis and a list of distress sentences
- phonemic level
  - not biased by orthography
  - word ending missing
  - slight decoding error
- low distance indicates proximity

What is the best threshold?

Distress detection results



ROC curve:

True positive = f(False negative)

Equal Error Rate: cutoff=0.7

Positive and negative test

Threshold = 0.7	Distress	Casual
d $\leq$ threshold	TP = 2472	FP = 374
d $>$ threshold	FN = 324	TH = 2632

recall = 88.4%
precision = 86.9%

F-measure = 87.2%

Conclusions and future work

Conclusions:

- Degradation of performances with the elderly voices
- Large variability of ASR performance inside each GIR group
- For a part of the elderly population, ASR must be adapted to each individual
- Correlation between performance and physical degradation

Future work:

- Experiments in real conditions in a smart home with CirdoX software
- Influence of the affects for automatic speech recognition

Speech Recognition of Aged Voice in the AAL Context: Detection of Distress Sentences

Thank you for your attention.

For Further Reading

M. Vacher, F. Portet, A. Fleury and N. Noury

Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges,

International Journal of E-Health and Medical Communications, **2**(1):35-54, January-March 2011.

M. Vacher, F. Portet, S. Rossato, F. Aman, C. Golanski and R. Dugheanu

Speech-based interaction in an AAL context

Gerontechnology,

11(2):310, 2012.

F. Portet, M. Vacher, C. Golanski, C. Roux, B. Meillon

Design and evaluation of a smart home voice interface for the elderly — Acceptability and objection aspects

Personal and Ubiquitous Computing

17(1):127–144, 2013.