Recognition of Distress Calls in Distant Speech Setting: a Preliminary Experiment in a Smart Home

Michel Vacher, Benjamin Lecouteux, Frédéric Aman, Solange Rossato, François Portet

Recognition of Distress Calls in Distant Speech Setting: a Preliminary Experiment in a Smart Home

Michel Vacher[1], *Benjamin Lecouteux*[2], *Frédéric Aman*[1], *Solange Rossato*[2], *François Portet*[2]

[1]CNRS, LIG, F-38000 Grenoble, France
[2]Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France
Michel.Vacher@imag.fr, Benjamin.Lecouteux@imag.fr, Frederic.Aman@imag.fr, Solange.Rossato@imag.fr, Francois.Portet@imag.fr

Introduction and context

Ageing of the population

- Life expectancy is growing up
  (≈ 22% more than 65 years old in 2050)

⟹
- Not enough places in special institutions
- 80% of people above 65 prefer to stay living at home

- Consequences of aging
  - Growing Isolation
  - Chronic and degenerative diseases (Alzheimer)
  - Reduced autonomy

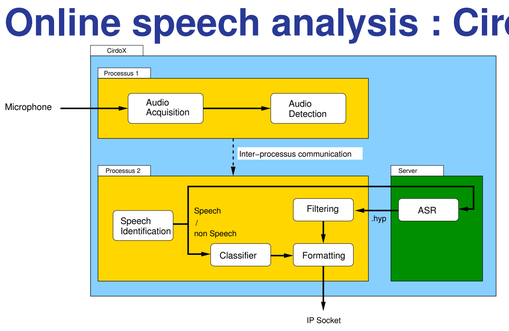
Smart Homes : A social issue
Adaptability according to the evolution of the person and of its needs : to help individuals retaining control of their environment

Speech use in Smart Home Projects

- Automatic speech recognition for elderly voices : VIPPERLA, PELLEGRINI
- Elderly people assistance aim but studies involving typical non-aged people : COMPANIONABLE, COMPANIONS, DIRHA
- Atypical voices (Alzheimer) : ALADIN, HOMESERVICE, PIPIN
- Vocal command system for home automation evaluated in a smart home by elderly and visually impaired people : SWEET-HOME
- Call for help by elderly people in distress case : This study : CIRDO
  - After a fall due to the carpet
  - In case of blocked hip when the person is sitten on the sofa

Method for distress call recognition

Online speech analysis : CirdoX system



Automatic speech recognition : acoustic modeling

- Kaldi speech recognition tool-kit was chosen as ASR system
- SGMM shared parameters using both SWEET-HOME data (7h), Voix-détresse (28mn) and clean data (ESTER+REPERE 500h).

Recognition of distress calls :

phonetic distance from a hypothesis to a list of predefined distress calls.

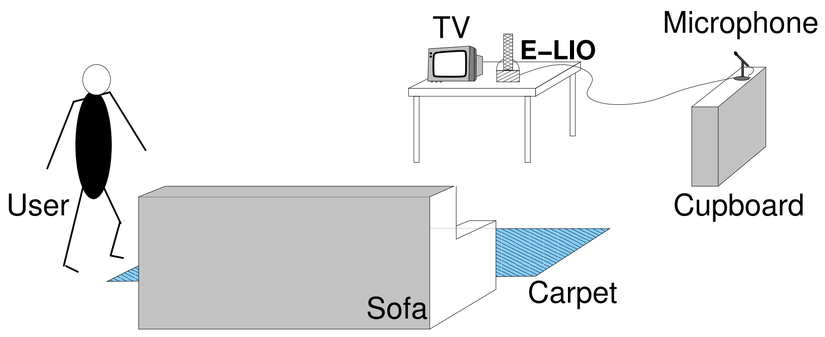
- Each ASR hypothesis H_i is phonetized, every voice commands T_j is aligned to H_i using Levenshtein distance.
- Deletion, insertion and substitution costs were computed empirically while the cumulative distance $\gamma(i,j)$ between H_j and T_i is given by :

$$\gamma(i,j) = d(T_i,H_j) + min\{\gamma(i-1,j-1),\gamma(i-1,j),\gamma(i,j-1)\}$$

- The decision to select or not a detected sentence is then taken according a detection threshold on the aligned symbol score (phonems) of each identified call.

Live experiment environment

Scenarios and experimental protocol



Distress calls, commands

Distress Sentence	Home Automation Command
Aïe aïe aïe ⋆	Appelle quelqu'un e-lio ⋆
Oh là ⋆	e-lio, appelle quelqu'un ⋆
Merde ⋆	e-lio tu peux appeler une ambulance
Je suis tombé ⋆	e-lio tu peux téléphoner au SAMU
Je peux pas me relever ⋆	e-lio, appelle du secours
Qu'est-ce qu'il m'arrive ⋆	e-lio appelle les secours
Aïe ! J'ai mal ⋆	e-lio appelle ma fille
Oh là ! Je saigne ! Je me suis blessé ⋆	e-lio appelle les secours

⋆ denotes a sentence identified during the sociological study by M.E. Bobillier Chaumont et al.

S. Bouaka et al., CIRDO : Smart companion for helping elderly to live at home for longer, IRBM, 35(2) :101-108

Scenarios, recorded corpus and off line experiments

Scenarios :
- 4 falls
- 1 blocked hip
- 2 "true-false" for video analysis

Simulator which hampered mobility for participant under 60 years old

Spk.	Age	Sex	Nb. of interjections or short sentences	
			All	*Distress*
S01	30	M	22	14
S02	-	-	-	
S03	24	F	16	15
S04	83	F	65	53
S05	29	M	24	21
S06	64	F	23	19
S07	61	M	23	21
S08	44	M	25	15
S09	16	M	32	21
S10	16	M	19	15
S11	52	M	12	12
S12	28	M	15	12
S13	66	M	24	21
S14	52	F	23	21
S15	23	M	20	19
S16	40	F	29	27
S17	40	F	24	21
S18	25	F	17	14
All	40.76		413	341

- Run on the Cirdo-set corpus
- SGMM as acoutic model.
- *generic LM* estimated from French newswire collected in the Gigaword corpus, 13K words
- Interpolated with *specialized LM (90%* (sentences used during the corpus collection)
- $CER = \frac{\text{Number of missed calls}}{\text{Number of calls}}$

	WER (%)				WER (%)		
Spk.	All	Distress	CER	Spk.	All	Distress	CER (%)
S01	45.0	39.1	27.8	S11	21.3	17.0	16.7
S03	41.4	44.4	40.0	S12	30.8	25.0	25.0
S04	51.9	49.6	34.0	S13	45.9	43.6	23.8
S05	19.1	15.4	14.3	S14	67.0	54.8	50.0
S06	39.2	34.3	26.3	S15	21.5	19.5	5.3
S07	21.2	20.3	28.6	S16	14.9	11.76	7.4
S08	61.8	50.8	20.0	S17	21.4	22.4	19.0
S09	49.4	41.2	33.3	S18	57.7	44.9	71.4
S10	24.5	22.4	14.3	All	39.3	34.0	26.8

TABLE: Word and Call Error Rate for each participant

Discussion and conclusion

- Global value of CER : 26.8% and 74.2% of calls correctly recognized
- If the system did not identify the first distress call because the person's voice is altered by the stress
- Study is focused on the framework of ASR applications in smart homes, that is in distant speech conditions and especially in realistic conditions very different from those of corpus recording when the speaker is reading a text.
- We presented the Cirdo-set corpus. The WER obtained at the output of the dedicated ASR was 36.3% for the distress calls.
- Thanks to a filtering of the ASR hypothesis at phonetic level, more than 70% of the calls were detected.
- Obtained results are not sufficient to allow the system use in real conditions and two research ideas can be considered.
- Speech recognition performances may be improved thanks to acoustic models adapted to expressive speech.
- It may be possible to recognize the repetition, at regular intervals, of speech events that are phonetically similar.

Contacts: [2] **Benjamin.Lecouteux@imag.fr**