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

To cite this version:
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. SLPAT, 2015, Dresde, Germany. hal-02088898

HAL Id: hal-02088898
https://hal.science/hal-02088898
Submitted on 3 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Recognition of Distress Calls in Distant Speech Setting: a Preliminary Experiment in a Smart Home

Michel Vacher¹, Benjamin Lecouteux², Frédéric Aman¹, Solange Rossato², François Portet²

¹CNRS, LIG, F-38000 Grenoble, France
²Univ. 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 2000)
- Not enough places in special institutions
- 85% of people above 65 prefer to stay living at home
- Consequences of ageing
  - 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

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-détéresse (28mn) and clean data (ESTER-REPERE 500h).

Live experiment environment

Scenarios and experimental protocol

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

- Run on the Cirdo-set corpus
- SGMM as acoustic 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)
- \( \text{WER} = \frac{\text{Number of missed calls}}{\text{Number of calls}} \)

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.

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, HOME SERVICE, PIPI
- 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

Recognition of distress calls :
phonetic distance from a hypothesis to a list of predefined distress calls.
- Each ASR hypothesis \( H \) is phonetized, every voice commands \( T_j \) is aligned to \( H \) using Levenshtein distance.
- Deletion, insertion and substitution costs were computed empirically while the cumulative distance \( \gamma(i,j) \) between \( H \) and \( T_j \) is given by:
  \[
  \gamma(i,j) = d(T_j,H) + \min\{\gamma(i−1,j−1),\gamma(i−1,j),\gamma(i,j−1)\}
  \]
- The decision to select or not a detected sentence is then taken according a detection threshold on the aligned symbol score (phonemes) of each identified call.

\[
\begin{array}{cccc}
\text{Distress Sentence} & \text{Home Automation Command} \\
\text{Aie aie aie} & \text{Appele quelqu’un} \{ \text{e-lio} \} \\
\text{Oh là là} & \text{e-lio, appelle quelqu’un} \\
\text{Merdès} & \text{e-lio tu peux appeler une ambulance} \\
\text{Je suis tombé} & \text{e-lio tu peux appeler au SAMU} \\
\text{Je peux pas me relever} & \text{e-lio appelle du secours} \\
\text{Ou est-ce qu’il m’arrive} & \text{e-lio appel les secours} \\
\text{Aie! J’ai mal} & \text{e-lio appelle ma fille} \\
\text{Oh là ! Je saigne ! Je me suis blessé} & \text{e-lio appelle les secours}
\end{array}
\]
- \( * \) 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

\[
\text{WER} = \frac{\text{Number of missed words}}{\text{Total number of words}}
\]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>WER (%)</th>
<th>Spk.</th>
<th>M</th>
<th>Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021</td>
<td>45.0</td>
<td>39.1</td>
<td>27.8</td>
<td>21.3</td>
</tr>
<tr>
<td>2022</td>
<td>41.4</td>
<td>44.0</td>
<td>20.0</td>
<td>30.9</td>
</tr>
<tr>
<td>2023</td>
<td>51.9</td>
<td>49.6</td>
<td>34.0</td>
<td>45.9</td>
</tr>
<tr>
<td>2024</td>
<td>19.1</td>
<td>15.4</td>
<td>14.3</td>
<td>70.7</td>
</tr>
<tr>
<td>2025</td>
<td>39.2</td>
<td>34.3</td>
<td>26.3</td>
<td>21.5</td>
</tr>
<tr>
<td>2026</td>
<td>21.2</td>
<td>20.3</td>
<td>28.6</td>
<td>14.9</td>
</tr>
<tr>
<td>2027</td>
<td>61.8</td>
<td>50.8</td>
<td>20.0</td>
<td>21.7</td>
</tr>
<tr>
<td>2028</td>
<td>45.4</td>
<td>41.2</td>
<td>33.3</td>
<td>57.7</td>
</tr>
<tr>
<td>2029</td>
<td>24.5</td>
<td>22.4</td>
<td>14.3</td>
<td>22.3</td>
</tr>
</tbody>
</table>

TABLE: WER and Call Error Rate for each participant