A Study on the Lombard Effect in Telepresence Robotics

Ambre Davat, GIPSA-lab / LIG, CNRS, Univ. Grenoble Alpes, Grenoble INP
Gang Feng, GIPSA-lab, CNRS, Univ. Grenoble Alpes, Grenoble INP
Véronique Aubergé, LIG, CNRS, Univ. Grenoble Alpes, Grenoble INP
What is a telepresence robot?  

>> a mobile robot piloted remotely to interact with other people

RobAIR Social Touch

- Designed at the FabMSTIC (fablab of University Grenoble Alpes) & Domus (LIG’s Living Lab)
- A tailor-made robot
- Experimentation out of the laboratory

Specific issue: connection between the pilot and the robot’s environment
Augmented telecommunication

>> better than traditional videoconferencing system

remote space

local space

Embodied pilot
=> physical presence
in the local environment
Augmented telecommunication

>> better than traditional videoconferencing system

BUT not perfect

remote space

Incomplete immersion
=> delay > 150 ms
=> no head tracking

local space

Robot ≠ clone of the pilot
voice not faithfully transmitted

(Počta & Komperda, 2016)
Unfaithful transmission of the voice

Ex: voice intensity

Risk of misunderstanding and break of social rules
Solution

>> binaural acoustic immersion
>> calibration of the audio channel

- sound intensity heard by the pilot = sound intensity in the local space
- voice intensity of the robot = voice intensity of the pilot

proprioceptive & interproprioceptive control of voice intensity
Problematic

Does **noise** in telepresence have different effects than in face-to-face interaction?
The Lombard Effect

= automatic regulation to noise
(Lombard, 1911)

• Affects **intensity**, pitch, spectral energy, speech rate, articulatory movements... (Cooke et al., 2014)

• Common to other animals (Brumm & Zollinger, 2011)

• Stronger during interactive tasks (Garnier et al. 2010)

• Hard to consciously inhibit (Pick et al. 1989)
Noise sources in Telepresence Robotics

Two cases:

=> Noise in the **local space**: both subjects can hear it.

=> Noise in the **remote space**: only the pilot can hear it.
Noise sources in Telepresence Robotics

**Two cases:**

=> Noise in the **local space**: both subjects can hear it.

=> Noise in the **remote space**: only the pilot can hear it.
Our Approach

An experiment to study how two interlocutors using a telepresence robot adapt to different noise sources

• measuring the Lombard Effect of both subjects

Standard case: both can hear the noise
Problemsatic case: only one of them can hear the noise

• individual or system-wide regulation to noise?
Test conditions

Subject R

remote space

Experimenter

Subject L

local space
Test conditions

A - quiet in both rooms
Test conditions

A - quiet in both rooms

B - noise in the remote space
Test conditions

A - quiet in both rooms
B - noise in the remote space
C - « virtual noise » in the local space
Test conditions

A - quiet in both rooms
B - noise in the remote space
C - « virtual noise » in the local space
D - noise in the local space
Test conditions

A - quiet in both rooms
B - noise in the remote space
C - « virtual noise » in the local space
D - noise in the local space

≈ 55 dB(A)
Expectations

Increase of voice intensity:
- if the subject can hear the noise
- when the other subject speaks louder
  = entainment effect (Székely et al., 2015)
Pretext task: test of the robot’s interface

**Subjects L**
Asks a list of questions & instructions

Ex: « What is the color of the sky? »
« Change the camera. »

**Subjects R**
Answers

The sky is blue.

Noise apparently *incidental* (B: coffee machine & C/D: drill)
Data

≈ 130 questions
+ 13 instructions for each session (≈ 20 min)

14 pairs of subject
(25/28 were native French speakers)

4 recordings:
- Voice of both subjects (headworn microphones) => Measures
- Microphone of the laptop of subject R => Noise B
- Monitoring signal of subject R => Noise C
- Pair of microphones of the robot => Noise D

<table>
<thead>
<tr>
<th>Condition</th>
<th>R</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1521</td>
<td>1322</td>
</tr>
<tr>
<td>B</td>
<td>182</td>
<td>127</td>
</tr>
<tr>
<td>C</td>
<td>161</td>
<td>139</td>
</tr>
<tr>
<td>D</td>
<td>163</td>
<td>152</td>
</tr>
</tbody>
</table>
Results

$max intensity \sim noise\ condition + (1|keyword) + (1|subject)$

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subjects R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Maximum intensity (dB)</td>
<td>64.61</td>
</tr>
<tr>
<td>Standard error (dB)</td>
<td>1.11</td>
</tr>
<tr>
<td>Number of extracts</td>
<td>1263</td>
</tr>
<tr>
<td>Statistical significance</td>
<td>$\chi^2(1) = 98.47 ; p &lt; 2.2 \times 10^{-16}$</td>
</tr>
</tbody>
</table>

Linear mixed effects model
(Winter Bodo 2013)
Results

**Subjects R (remote pilots)**

- Augmentation of voice intensity in each noise condition

  - **B** (Direct noise of a coffee machine) > \{ C, D \} (Drill noise heard with headphones)
  
  - **D** (R & L hear the noise) > C (Only R hears the noise)

=> entrainment effect?
Results

Subjects R (remote pilots)

- Augmentation of voice intensity in each noise condition

- B (Direct noise of a coffee machine) > { C, D } (Drill noise heard with headphones)

- D (R & L hear the noise) > C (Only R hears the noise)

<table>
<thead>
<tr>
<th>Subjects L</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>69.19</td>
<td>-0.06</td>
<td>-0.30</td>
<td>+2.24</td>
</tr>
<tr>
<td>DR</td>
<td>1.86</td>
<td>0.21</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>DR</td>
<td>1125</td>
<td>118</td>
<td>129</td>
<td>143</td>
</tr>
</tbody>
</table>

\[ \chi^2(1) = 132.59; p < 2.2 \times 10^{-16} \]
Results

Subjects R (remote pilots)
• Augmentation of voice intensity in each noise condition
• B (Direct noise of a coffee machine) > { C, D } (Drill noise heard with headphones)
• D (R & L hear the noise) > C (Only R hears the noise)

Subjects L (local interlocutors)
• Augmentation of voice intensity only in condition D

=> entrainment effect
Conclusion

Lombard effect:
• in every noisy condition
• smaller than in (Winkworth & Davis 1997) (+1 to 3 dB instead of +10 dB), but significative

Perspective

• Using different noises (duration, type, intensity)
• Check if people are able to perceive these small intensity variations
  => If so, how do they understand them?
Thank you for your attention!
Living Lab Domus

Office

Office

Intelligent apartment

Corridor

Sound-proof room

Control room

Experimental platform
Results

14 pairs of subjects

R model: \( \text{max intensity} \sim \text{noise condition} + (1|\text{keyword}) + (1|\text{subject}) \)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subjects R</th>
<th></th>
<th></th>
<th></th>
<th>Subjects L</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum intensity (dB)</td>
<td>64.61</td>
<td>+ 2.98</td>
<td>+ 1.19</td>
<td>+ 2.38</td>
<td>69.19</td>
<td>- 0.06</td>
<td>- 0.30</td>
<td>+ 2.24</td>
</tr>
<tr>
<td>Standard error (dB)</td>
<td>1.11</td>
<td>0.36</td>
<td>0.37</td>
<td>0.36</td>
<td>1.86</td>
<td>0.21</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>Number of extracts</td>
<td>1263</td>
<td>160</td>
<td>146</td>
<td>150</td>
<td>1125</td>
<td>118</td>
<td>129</td>
<td>143</td>
</tr>
<tr>
<td>Statistical significance</td>
<td>( \chi^2(1) = 98.47; p &lt; 2.2 \times 10^{-16} )</td>
<td></td>
<td></td>
<td></td>
<td>( \chi^2(1) = 132.59; p &lt; 2.2 \times 10^{-16} )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>