### M.1. General Information

**Data included:**

- Audio (see section M.6.)
- Image (see section M.7.)
- Video (see section M.8.)

**Language(s):**

- Language dependent
- Language independent

### M.2. Recording Information – Humans

**Sex and number of humans:**

- Male Number:
- Female Number:
- Imposters Number:
- Synthetic Number:

- Total number:

**Number of humans visible in the same frame:**

**Age class:** (please indicate number of speakers and age)

- Children I ( __ to __ )
- Children II ( __ to __ )
- Teenagers I ( __ to __ )
- Teenagers II ( __ to __ )
- Adults I ( __ to __ )
- Adults II ( __ to __ )
- Adults III ( __ to __ )
- Adults IV ( __ to __ )
- Elderly (over 60)
- Age unknown
- Other distribution:

**Origin:**

- Native
- Non native
- Unknown

**Geographic distribution:**

- Total number of regions:
- Percentage per region:

**Regions included:**

**Information included about:**

- Place of living
- Place of birth
- Dialect/accent

**Additional speaker information included:**

- Speaking/hearing impairments
- Height
- Weight
- Smoking habits
- Trained speakers
- Education level
- Profession
- Place of (secondary) education
- Other:

**Comments:**
### M.3. RECORDING INFORMATION – RESOURCE

<table>
<thead>
<tr>
<th>Human body parts visible in the resource:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ None</td>
<td>□ Whole body</td>
</tr>
<tr>
<td>□ Face</td>
<td>□ Mouth</td>
</tr>
<tr>
<td>□ Hands</td>
<td>□ Legs</td>
</tr>
<tr>
<td>□ Other</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distractors visible in the resource:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ None</td>
<td>□ Hat</td>
</tr>
<tr>
<td>□ Watch</td>
<td>□ Scarf</td>
</tr>
<tr>
<td>□ Microphone</td>
<td>□ Markers</td>
</tr>
<tr>
<td>□ Other</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interactive media visible/audible in the resource:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ None</td>
<td>□ Graphical screen</td>
</tr>
<tr>
<td>□ Tactile screen</td>
<td>□ Data glove</td>
</tr>
<tr>
<td>□ Desktop PC</td>
<td>□ Laptop</td>
</tr>
<tr>
<td>□ Other</td>
<td>□</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotated modalities in the resource: (for details see section M.4.)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ None</td>
<td>□ Speech</td>
</tr>
<tr>
<td>□ Gaze/Eye movements</td>
<td>□ Facial expressions</td>
</tr>
<tr>
<td>□ Head movements</td>
<td>□ Body movements</td>
</tr>
<tr>
<td>□ Other</td>
<td>□</td>
</tr>
</tbody>
</table>

Total number of annotated modalities:

<table>
<thead>
<tr>
<th>Other modalities available/visible but not annotated in the resource: (for details see section M.4.)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ None</td>
<td>□ Speech</td>
</tr>
<tr>
<td>□ Gaze/Eye movements</td>
<td>□ Facial expressions</td>
</tr>
<tr>
<td>□ Head movements</td>
<td>□ Body movements</td>
</tr>
<tr>
<td>□ Other</td>
<td>□</td>
</tr>
</tbody>
</table>

Total number of modalities (not annotated):

<table>
<thead>
<tr>
<th>Scene – Illumination:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Daylight</td>
<td>□ Single source</td>
</tr>
<tr>
<td>□ Fix</td>
<td>□ Variable</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scene – Backgrounds:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Plain</td>
<td>□ Complex</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sessions:</td>
<td></td>
</tr>
<tr>
<td>Number of poses per subject:</td>
<td></td>
</tr>
</tbody>
</table>
### M.4. Modalities – Detailed Information

#### Facial expressions:
- (6 universal expressions of emotion)
- Anger
- Happiness
- Other:
- Disgust
- Sadness
- Fear
- Surprise

#### Head movements:
- Rotation
- Other:
- Inclination forward/backward
- Inclination sideward

#### Face views:
- Frontal
- Other:
- Profile
- Other:
- Total number of face views per subject:

#### Gestures:
- Interactional
- Non-interactional
- Non identified
- Other:
- Emotional

#### Gaze/Eye movements:
- Saccades
- Pursuit motion
- Horizontal
- Vertical
- Other:
- Convergence

#### Hand manipulation of objects:
- Direct manipulation
- Joining/splitting objects
- Other:
- Indirect manipulation
- Changing object position
- Modifying objects

#### Body movements:
- Upper body
- Arms
- Other:
- Lower body
- Legs
- Whole body

### M.5. Application Information

#### Authentication:
- Face verification
- Other:
- Speech verification
- User authentication

#### Recognition:
- Face recognition
- Other:
- Automatic speech recognition
- Automatic person recognition
- Expression recognition
- Other:
- Automatic person recognition

#### Analysis:
- Lip tracking
- Other:
- Speech/lips correlation
- Other:

#### Synthesis:
- Talking heads
- Multimedia development
- Other:
- Avatars
- Humanoid agents

#### Control:
- Voice control
- Other:
- Speech assisted video
- Other:

#### Miscellaneous:
- Information retrieval
- Other:

#### Application areas:
- Education/Training
- Banking
- Research
- Tourism
- Entertainment
- Other:
**M.6. TECHNICAL INFORMATION – AUDIO**

**Signal encoding:**
- [ ] A-law
- [ ] μ-law
- [ ] Linear PCM
- [ ] Other:

**File format:**
- [ ] AIFF
- [ ] Wav
- [ ] Without header
- [ ] SAM
- [ ] NIST/Sphere
- [ ] Au
- [ ] Other:

**Sampling rate:**
- [ ] 8 kHz
- [ ] 16 kHz
- [ ] 32 kHz
- [ ] 44.1 kHz
- [ ] 48 kHz
- [ ] Other:

**Quantisation:**
- [ ] 8 bit
- [ ] 16 bit
- [ ] 32 bit
- [ ] 64 bit
- [ ] Other:

**Byte order:**
- [ ] Lo-hi (Intel)
- [ ] Hi-lo (Motorola)

**Data format:**
- [ ] Signed integer
- [ ] Unsigned integer
- [ ] Other:

**Amount of data:**
Size (Mb, Gb, etc) or duration (minutes, hours, etc):

**Compression:**
- [ ] None
- [ ] Shorten
- [ ] Other:

**Number of recording channels:**
- [ ] 1 (mono)
- [ ] 2 (stereo)
- [ ] 3
- [ ] 4
- [ ] 8
- [ ] Other:

**Annotation standard:**
- [ ] SAM
- [ ] SGML
- [ ] XML
- [ ] NIST/LDC
- [ ] Other:

**Sound quality measures included:**
- [ ] SNR
- [ ] Cross talk
- [ ] Clipping rate
- [ ] Background noise
- [ ] Other:

**Tools used for measuring sound quality:**

**Speech content:**
<table>
<thead>
<tr>
<th>Speech content</th>
<th>No. of items:</th>
<th>No. of items:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application words</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Concatenated words</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Isolated words</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Syllables</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>VCV sequences</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Continuous sentences</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Phonetically rich sentences</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Phonetically balanced sentences</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>Digit-set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isolated digits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Natural numbers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Money amounts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit card numbers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone numbers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes/no questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### M.7. Technical Information – Image

**Resolution in pixels:**

**Color components:**
- [ ] RGB
- [ ] CMYK
- [ ] 4:2:2
- [ ] Other:

**Color depth:**
- [ ] 8 bits
- [ ] 16 bits
- [ ] 24 bits
- [ ] Other:

**File format:**
- [ ] JPG
- [ ] GIF
- [ ] TIFF
- [ ] BMP
- [ ] EPS
- [ ] CIF
- [ ] PPM
- [ ] Other:

**Amount of data:**
- Size (Mb, Gb, etc):
- Duration (minutes, hours, etc):

**Compression:**
- [ ] None
- [ ] Zip
- [ ] Other:

**Compression ratio:**

### M.8. Technical Information – Video

- [ ] Synchronized audio (see section M.6.)

**Resolution in pixels:**

**Color components:**
- [ ] RGB
- [ ] CMYK
- [ ] 4:2:2
- [ ] Other:

**Color depth:**
- [ ] 8 bits
- [ ] 16 bits
- [ ] 24 bits
- [ ] Other:

**Frame rate:**
- [ ] <25 frames/sec.
- [ ] 25 frames/sec.
- [ ] 30 frames/sec.
- [ ] 50 frames/sec.
- [ ] 60 frames/sec.
- [ ] Other:

**File format:**
- [ ] MOV
- [ ] AVI
- [ ] MPEG
- [ ] Quicktime
- [ ] SGI
- [ ] Other:

**Amount of data:**
- Size (Mb, Gb, etc):
- Duration (minutes, hours, etc):

**Compression:**
- [ ] None
- [ ] Zip
- [ ] Other:

**Compression ratio:**
### M.9. TECHNICAL INFORMATION – MODELING

<table>
<thead>
<tr>
<th>Models:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ 2D</td>
<td>☐ 3D</td>
<td>☐ Other:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File formats:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ VRML</td>
<td>☐ Other:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms used:</th>
<th></th>
</tr>
</thead>
</table>

### M.10. FURTHER COMMENTS

---

ELRA
55-57 rue Brillat Savarin
F-75013 PARIS, FRANCE
Tel +33 1 43 13 33 33 - Fax +33 1 43 13 33 30
http://www.elra.info