Evaluation and Context for In-car Speech Systems for Older Adults

Philipp Schneider
Seminar Personalizing the User Experience
Supervisor: Michael Feld
29.06.2012
Teaser - Motivation
Motivation (1)

Many car voices in the movie history
Motivation (2)

• What voice is best in an in-car speech system?
• Should be:
  ▫ Receptive
  ▫ Credible
  ▫ Convincing

Next street on the right in 600 m
Outline

- Motivation
- Introduction
- Voice perception
- Experiments
- Results
- Conclusion
Introduction (1)

Which driver prefers which voice?
Introduction (2)

Population
- People in UK over 65
- People in UK under 65

Population
- People in UK over 65 (2031)
- People in UK under 65 (2031)

special needs for older adults
Introduction (3)

• In-car information systems -> common
  ▫ Some rely on speech-based interactions
    ➞ many questions

One voice for all types of information content?
One voice per information category?
One voice per age group of drivers?
Introduction (4)

• This paper: optimum age of voice for
  ▫ 18-25 years
  ▫ 55 years and older

Experiments in laboratory setting

Experiments in driving simulation
Voice perception (1)

• Humans can detect characteristics in voice
  ▫ Often use that ability subconsciously
    • People like consistency in voice
    • Voice influences perception of content
    • People prefer voice characteristics matching content

• How does one detect characteristics?
Voice perception (2)

- Christian Wolff's Seminar about individuality
  - Pitch
  - Volume
  - Rate
  - Tone
  - Jitter
  - Intensity
  - Shimmer
- Voice changes with age
Voice perception (3)

- What do we perceive?
- Possible to estimate
  - Age
  - Likeability
  - Gender
  - Emotion
  - Stress

What is the impact on people’s perception of information, system and car when the characteristic of the voice is mismatched with the content of the information and what can you do with that abilities?
Voice perception (4)

- Examples female voice
  - Apple’s Siri
  - Some cockpits
  - Automated tech support line
- Examples male voice
  - Navigation system
Voice perception (5)

• What can one do with variation of these characteristics?
• Characteristics in voice can influence:
  • Attention
  • Performance
  • Judgement
  • Risk-taking

Calm down!
Rationale for experiments

• Nature and quality of voice important
• speech-based in-car information messages:
  ▫ Able to instil confidence
  ▫ Compensate for memory loss
  ▫ Help to make contextually relevant suggestions
  ▫ Provide contextually relevant advice
• User‘s perception of speech-based support system as important as it‘s functionality
Laboratory experiments (1)

- Two voices: 20 & 73 year old woman
- Two listener groups: 18-25 years & 55+ years

3 questions

- Are there differences in the emotional coloring, the perceived trust, and credibility between the two voices?
- Is there any difference in the voice quality between the two voices?
- Similarity – what are the perceived age, background and attitude of the persons speaking?
Laboratory experiments (2)

• 26 short voice prompts, selectable and playable at any time within 30 minutes

• Examples:
  • There is thick fog ahead.
  • Beware of cyclists ahead.
  • The current speed limit is 60 miles an hour.
  • There is an accident ahead, turn right to avoid it.

• 12 participants – 6 male, 6 female – 6 over 55 years, 6 between 18 and 25 years
Laboratory experiments (3)

- Positive and negative emotional Coloring of Voice
  - 10-point scale (1=very bad described, 10=very well described)
  - Examples: happy, amused, mad, aggressive
- Trust of voice
  - Contrasting adjectives on ends of a 10-point scale
- Credibility of Voice
  - Contrasting adjectives on ends of a 10-point scale
Laboratory experiments (4)

- Quality of voice
  - Contrasting adjectives on ends of a 10-point scale
  - Participants should judge the age
- Homophily – Similarity
  - Contrasting statements on ends of a 10-point scale
  - Three indices
    ▫ Attitudinal similarity
    ▫ Behavioral similarity
    ▫ Similarity
Laboratory experiments (5)

- Emotional coloring of voice
  - Age groups
    - Older adults rated both voices significantly more positive
    - No significant difference in rating of negative emotional coloring
  - Two voices
    - No significant difference in positive and negative emotional coloring
Laboratory experiments (6)

- Trust of voice
  - Age groups
    - No significant difference in rating of trustworthiness
  - Two voices
    - Older voice significant more trustworthy than the younger voice
Laboratory experiments (7)

• Credibility of voice
  ▫ Age groups
    • No significant difference in rating of credibility in most parts
  ▫ Two voices
    • Older adult voice more authoritative and qualified
    • Young voice more dynamism
    • No difference in character
Laboratory experiments (8)

- Quality of Voice
  - Age groups
    - No significant difference in rating of clarity
  - Two voices
    - Older adult voice perceived to have more clarity than the young voice
Laboratory experiments (9)

- Homophily - Similarity
  - Age groups
    - Older group feeling more similar to older voice
    - Young group perceived differences on all similarity measures
  - Two voices
    - No significant difference in perceived similarity rating
Laboratory experiments (10)

Is there any difference in the trustworthiness between the two voices? **YES**

Is there any difference in the quality between the two voices? **NO**

Is there any difference in the emotional coloring between the two voices?

What are the perceived age of the person speaking?

Answer: 21/63 years
Correct: 20/73 years
Comparison with driving simulator results

• Expect: better driving with old voice
• But: Older adults drive significantly better with a young voice
• How is it explainable?

Perceived similarity:
Older adults aware of their declining physical and attention abilities
Young voice is associated with better abilities like vision and reflexes
Source credibility results (1)

Clear difference between younger and older adults

Source credibility of the two voices according to age of participants – laboratory experiment
Source credibility results (2)

Source credibility of the two voices according to age of participants - driving simulator

Source credibility of the young adult voice – driving simulator
Conclusion

- Older Adults prefer younger voice while driving
- Results of laboratory experiments can be perceived differently in reality/driving simulator
- Judgment of older adults was more affected by change in context
- Younger people appear to be more flexible

Different user groups – different needs!
Summary

Introduction (1)

Voice perception (4)
- Can estimate
  - Age
  - Likeability
  - Gender
- Examples female voice
  - Apple's Siri
  - Cockpit
  - Automated tech support

Laboratory experiments (1)
- Two voices: 20 & 73 year old woman
- Two listener groups: 18-25 years & 50+
- 3 questions
  - Are there differences in the emotional coloring, the perceived trust, and credibility between the two voices?
  - Is there any difference in the voice quality between the two voices?

Source credibility results (1)
- Clear difference between younger and older adults
- Source credibility of the two voices according to age of participants – laboratory experiment
Questions?
# Tables (1)

<table>
<thead>
<tr>
<th></th>
<th>Young Voice</th>
<th></th>
<th>Older Adult Voice</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age group</td>
<td>Age group</td>
<td>Age group</td>
<td>Age group</td>
</tr>
<tr>
<td></td>
<td>18-25</td>
<td>55 and over</td>
<td>18-25</td>
<td>55 and over</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Positive emotional coloring</td>
<td>35.0</td>
<td>6.6</td>
<td>54.8</td>
<td>6.7</td>
</tr>
<tr>
<td>Trust of voice</td>
<td>28.2</td>
<td>2.3</td>
<td>28.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Character of voice (credibility)</td>
<td>18.3</td>
<td>3.2</td>
<td>18.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Qualification of voice (credibility)</td>
<td>14.6</td>
<td>2.4</td>
<td>17.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Dynamism of voice (credibility)</td>
<td>29.9</td>
<td>4.7</td>
<td>24.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Clarity of voice (quality)</td>
<td>35.17</td>
<td>6.2</td>
<td>37.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Age of persons speaking</td>
<td>18.33</td>
<td>2.7</td>
<td>23.2</td>
<td>4.8</td>
</tr>
<tr>
<td>Attitudinal similarity</td>
<td>24.3</td>
<td>6.7</td>
<td>15.5</td>
<td>6.1</td>
</tr>
<tr>
<td>Behavioral similarity</td>
<td>23.3</td>
<td>5.2</td>
<td>23.3</td>
<td>2.0</td>
</tr>
<tr>
<td>Similarity (combined)</td>
<td>47.7</td>
<td>11.4</td>
<td>41.0</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Comparison: Rating of Voices by Age Group
## Tables (2)

<table>
<thead>
<tr>
<th></th>
<th>Young Voice</th>
<th></th>
<th>Older Adult Voice</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Positive emotional coloring</td>
<td>44.9</td>
<td>12.1</td>
<td>38.8</td>
<td>9.7</td>
</tr>
<tr>
<td>Negative emotional coloring</td>
<td>16.2</td>
<td>7.2</td>
<td>21.8</td>
<td>10</td>
</tr>
<tr>
<td>Trust of voice</td>
<td>28.5</td>
<td>2.8</td>
<td>47.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Authoritativness of voice (credibility)</td>
<td>29.2</td>
<td>6.4</td>
<td>36.2</td>
<td>5.6</td>
</tr>
<tr>
<td>Qualification of voice (credibility)</td>
<td>16.2</td>
<td>2.8</td>
<td>20.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Dynamism of voice (credibility)</td>
<td>27.4</td>
<td>4.4</td>
<td>22.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Clarity of voice (quality)</td>
<td>36.4</td>
<td>5.3</td>
<td>42.3</td>
<td>4.2</td>
</tr>
<tr>
<td>Similarity (combined)</td>
<td>44.3</td>
<td>9.0</td>
<td>36.2</td>
<td>18.2</td>
</tr>
</tbody>
</table>

Comparison: Young Voice and Older Adult Voice