The effect of learning on listening to ultra-fast speech

T. Nishimoto (The Univ. of Tokyo)
Y. Kariya and T. Watanabe
(Tokyo Woman’s Christian Univ.)
Objective

• Visually impaired persons
  – use PC with voice-enabling software
  – get information via synthesized voice
  – important to acquire information efficiently
  – **Ultra-fast speech** : they can practice

• This work : observe the learning effect
  – effect of word familiarity
  – intelligibility & mental workload
  • undesirable when mental workload increases
Prior work

- [Asakawa et. al. 2003, 2005]
  - fast speech by shortening the recorded human voices
  - by the skilled users of the screen reading software
  - The skilled listeners assessed
    - the appropriate speed as 19.5 morae/sec (recall 90%)
    - the ultimate speed as 25.0 morae/sec (recall 50%)

- [Nishimoto et. al. 2006, 2007]
  - TTS voices: 4-digit numbers (18-30 morae/sec)
  - experiments with young persons
    - significant learning effects
    - sustained for several weeks
  - age-related effects: intercept the learning
    - difficulty of recall digits in the correct order
Listening ultra-fast speech

- **Bottom-up process**
  - dominant when low-familiarity task
  - learning effect: less significant
  - mental workload: high

- **Top-down process**
  - lexicon is important
  - mental workload: low

- **Word familiarity**
  - awareness influences the learning effect
Estimation of mental workload

- NASA-TLX: weighted mean workload
  - Mental demand / Physical demand
  - Temporal pressure / Effort
  - Frustration level / Performance
- easy to use, reliably sensitive

explanation  
ranking  
evaluation

magnitude of factor (6-1)  
rest time (5min) before next task

values of 0-100 for 6 factors

WWL score
Preliminary experiment

- Results: significance in WWL & intelligibility – except the case between FH-S2 and FL-S1
  - speed: S1=original / S2=2.0 times faster
  - familiarity: FH / FL

Normalization of WWL

Transformed the ave. and SD of WWL of each subject to 50 and 10, respectively
Familiarity controlled speech

- FW03 [Amano2006]
  - audio data set for word intelligibility test
    - based on Japanese word familiarity database
  - 4 levels of familiarity
    - 4 morae words: accent controlled
    - a group consists of 50 words: phoneme balanced
  - Recorded speech
    - speed controlled, loudness level calibrated
- Stimuli: audio speed conversion was used
  - 4 times faster = approx. 20 morae / sec
  - highest (H) / lowest (L) familiarity groups were used
- Evaluation: number of correct morae
Experiment

- Subjects (N=59) : divided into 4 groups
  - PC with headphones for everyone
  - Input roman characters with key-board
- Each trial consists of 50 words
  - 16 seconds interval
  - word familiarity was not informed

<table>
<thead>
<tr>
<th>Group</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1 : L-L-L</td>
<td>FLS1</td>
<td>FLS2</td>
<td>FLS3</td>
</tr>
<tr>
<td>G2 : H-H-L</td>
<td>FHS1</td>
<td>FHS2</td>
<td>FLS3</td>
</tr>
<tr>
<td>G3 : L-L-H</td>
<td>FLS1</td>
<td>FLS2</td>
<td>FHS3</td>
</tr>
<tr>
<td>G4 : H-H-H</td>
<td>FHS1</td>
<td>FHS2</td>
<td>FHS3</td>
</tr>
</tbody>
</table>

S1, S2, S3 : different vocabulary (same condition)

- difficult to use mental lexicon
- if they are not aware of the change, they do not use mental lexicon
- if they are aware of high familiarity, they can use mental lexicon
Intelligibility T1-T2, T1-T3: **

G1: L-L-L (N=16)

Intelligibility & WWL T2-T3, T1-T3: **

Intelligibility T1-T2, T2-T3, T1-T3: **

WWL T2-T3: *

**: p < 0.05
*: 0.05 < p < 0.10 (t-test of two-tailed distributions)

No significant differences between T3 of G1-G2 and T3 of G3-G4, respectively

peak of learning effect at T2

not significant WWL reduction at T3

significant WWL reduction at T3

G3: L-L-H (N=14)

G2: H-H-L (N=15)

G4: H-H-H (N=14)

Intelligibility T2-T3, T1-T3: **

WWL

T1-T2: *

T2-T3, T1-T3: **

No significant differences between T3 of G1-G2 and T3 of G3-G4, respectively
Conclusion

• Learning reduces mental workload if lexicon is used
• Without using lexicon, learning effect is limited
• Awareness is the key
  – indications may change
• Future work:
  – longer period experiment
  – age-related effects
  – TTS applications

Acknowledgment: A. Segawa (TWCU) performed additional experiment.