Delay perception thresholds in human-computer interaction

fundamentals for CSCW-applications

Conference Paper

Author(s):
Zuberbühler, Hans-Jörg; Krueger, Helmut; Kündig, Albert

Publication date:
2003

Permanent link:
https://doi.org/10.3929/ethz-a-004606991

Delay Perception Thresholds in Human-Computer Interaction: Fundamentals for CSCW-Applications

Hans-Jörg ZUBERBÜHLER[+], Helmut KRUEGER[+] and Albert KÜNDIG*

[+]*Institute of Hygiene and Applied Physiology, Swiss Federal Institute of Technology Zurich, Switzerland*

**Computer Engineering and Networks Laboratory, Swiss Federal Institute of Technology Zurich, Switzerland*

Abstract. In order to optimise resource allocation in telecommunication networks, it is important to have profound knowledge of the users' subjective perceptions of quality. By means of psychophysical methodology we examined the perception thresholds of a Quality of Service (QoS) parameter represented by the overall delay between sending information and receiving answer when interacting with a computer. The thresholds are at a delay between vocal input and visual response of 115 (± 23) ms, and at a delay between click input and visual response of 78 (± 14) ms (numbers in brackets stand for the 95% confidence intervals). These two threshold values differ significantly (p=0.01) from each other, whereas age and gender of the subjects do not exert significant influence on the location of the thresholds.

1. Introduction

Packet-switching technology (e.g. IP or ATM) is state-of-the-art in telecommunication networks. Compared to circuit-switching (e.g. ISDN) it increases network efficiency but – on the other hand – it decreases network reliability. That is, considerable variation in error rates, bit rates, and transmission delays may occur. These variations represent a challenge in the design and use of packet-switched networks, since users will be confronted with varying Quality of Service (QoS)[1]. To date, the majority of research on QoS is systems oriented, focusing on traffic analysis, scheduling, and routing. Relatively minor attention has been paid to user-level QoS issues (Bouch et al., 2000). It is not yet known how objective system quality relates to users' subjective perceptions of quality. This discrepancy is the driving force behind our work, aiming to examine the end-user’s perception of QoS, and trying to answer the question: ***How do network-induced impairments affect the interaction of two or more users of a telecommunication system****?* Such systems are typically pure videoconference systems (over fixed or mobile networks, e.g. UMTS) as well as systems enabling Computer Supported Cooperative Work (CSCW).

As a first step, we have answered the above mentioned question for the QoS-parameter *intermedia synchronisation* (Zuberbühler et al., 2002), which represents a relative delay parameter. The topic of this contribution is an absolute delay parameter represented by the overall delay between sending information and receiving answer. Therefore it is perceived only in dialogue settings. The below described experiments were performed to determine the smallest delays humans perceive when interacting with a computer. With these experiments we investigated two input modalities in which either a vocal or a tactile input triggered a computer generated visual feedback. The outcome of these experiments are considered as a prerequisite for the investigation of the more complex situation where humans interact with each other accomplishing realistic tasks.

---

[1] For the numerous definitions of Quality of Service (QoS) see (Fluckiger, 1995).

2. Method and Procedure

The problem at hand is one of threshold determination, a task for which psychophysical procedures have proven suitable. We used a two-alternative forced-choice (2AFC) task, and defined the threshold to be that delay yielding 75% of detection, since in 2AFC the rate of detection exclusively by guessing is 50%. In order to reach the threshold as fast as possible, we applied the adaptive procedure called *best-PEST*[2] (Pentland, 1980). Our experimental set up, including stimulus presentation, *best-PEST* algorithm, and data acquisition was implemented using *Macromedia Director*'s scripting language *Lingo*. Seven female and 17 male subjects (aged between 19 and 41, mean=25) were recruited for the two experiments. Each of the experiments lasted about 20 minutes. The subject had to complete the threshold procedure three times for both modalities, resulting in a total of 144 threshold estimations. The first threshold estimation in each condition was considered as practice and therefore excluded from further analysis.

2.1 Voice Trigger – Visual Response (VoiVis)

As visual stimulus, a black cipher with a height of 4 arc degrees appeared on a white background. The subjects were told to pronounce the displayed cipher. When the sound-level of the subject's voice exceeded 65.5 dBA the cipher disappeared either after a small, system-inherent delay, or after the delay calculated by *best-PEST*. The triggering sound level of 65.5 dBA was chosen as a result of pre-tests, showing that this value – on the one hand – is reached through normal speaking and – on the other hand – is high enough to avoid the disappearance of the cipher due to background noise.

2.2 Mouse Trigger – Visual Response (MouVis)

The visual stimulus consisted of a red square with a side length of 5 arc degrees on white background. The subjects were told to click into a white square, whereupon the square changed its colour to red, either immediately or after a delay calculated by *best-PEST*.

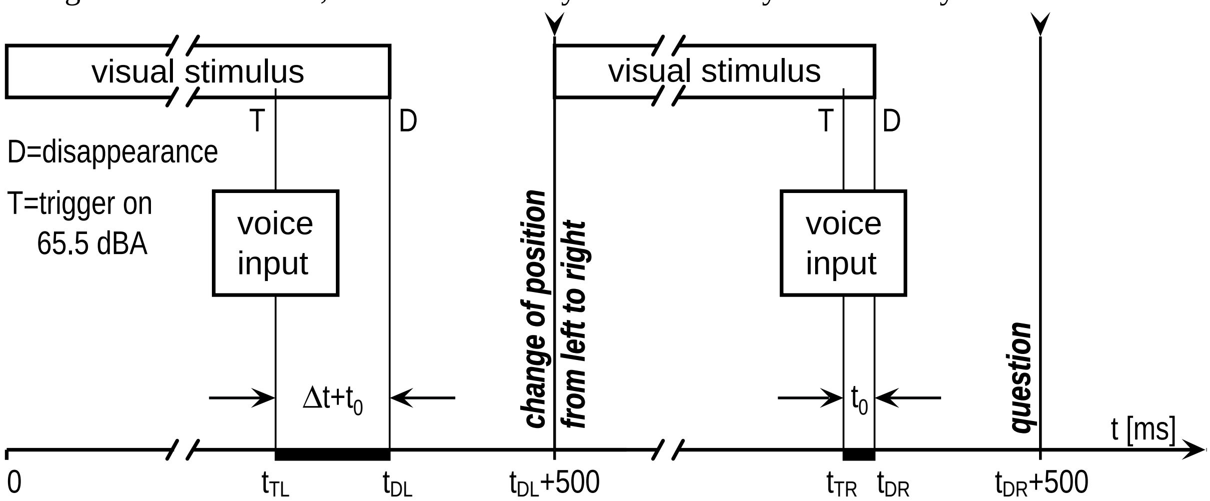

Figure 1: *Test sequence of one trial exemplified by the VoiVis-experiment. The duration of Δt is equal to the maximum likelihood of the threshold computed in the best-PEST procedure. The occurrence of Δt is randomly balanced between the left and the right side of the screen. t_0 is the average response latency of the microphone device. The question after each sequence was: "On which side did you perceive a delayed disappearance of the cipher?"*

---

[2] PEST is the acronym for ***P**arameter **E**stimation of **S**equential **T**rials.*

3. Results

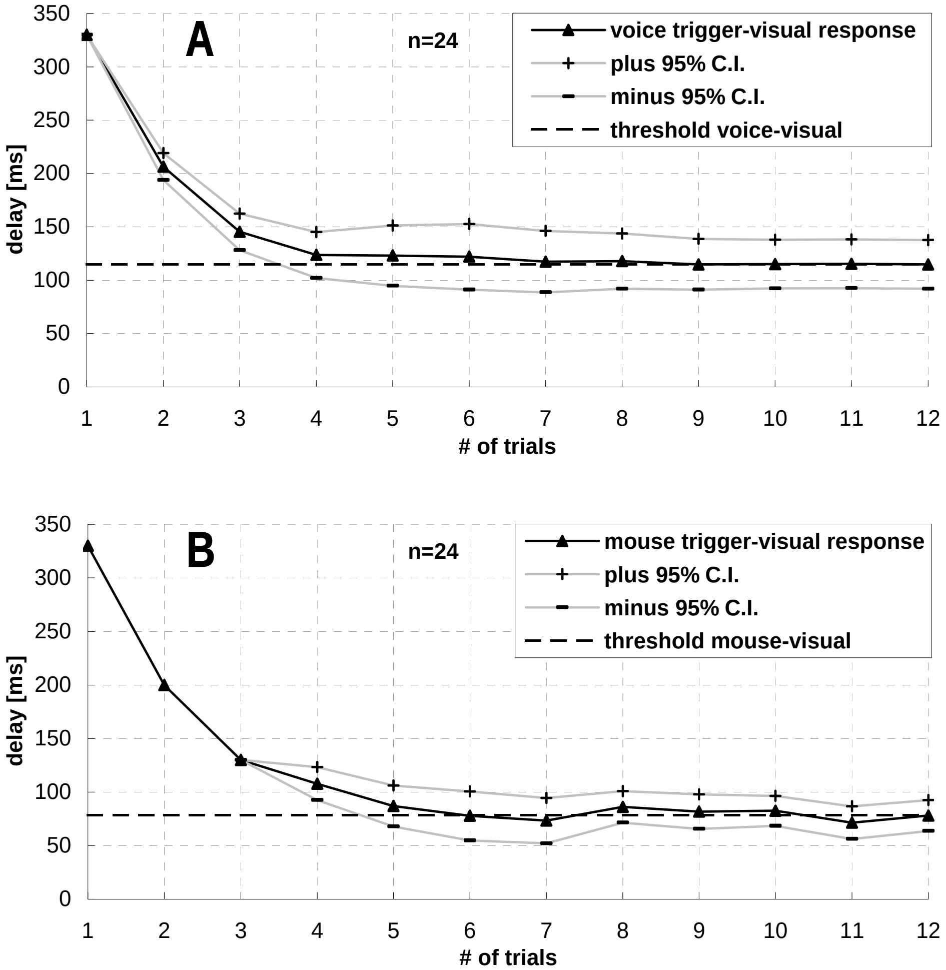

Figure 2: *Delays calculated by best-PEST for every trial. The curves represent the mean of all subjects. Continuous lines show the progression of the threshold convergence. Dashed lines indicate the final thresholds, grey lines indicate the 95% confidence intervals.*
***A**: Voice trigger – Visual response, **B**: Mouse trigger – Visual response.*

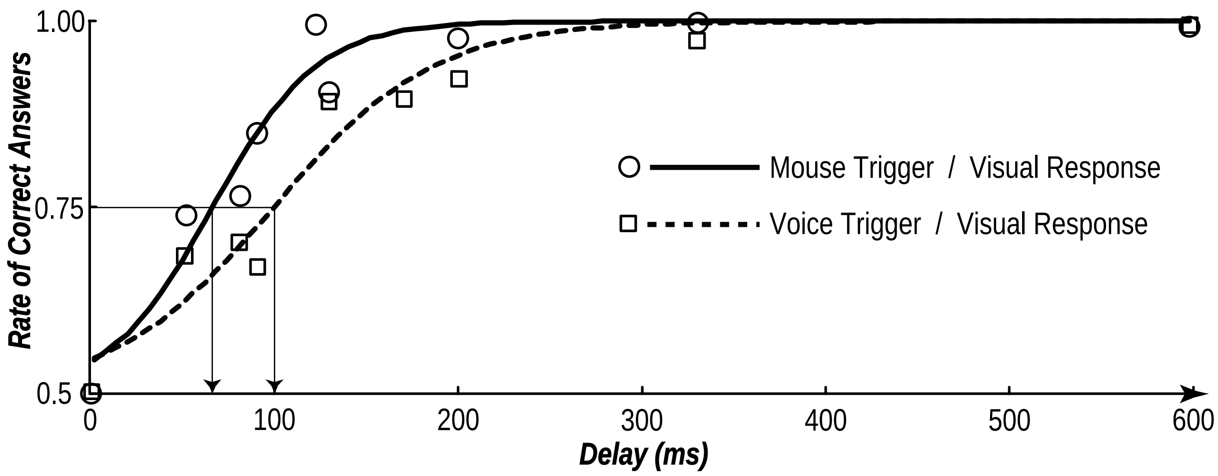

Figure 3: *Psychometric functions for both conditions. The obtained data are fitted with a logistic model (method of least squares). Since a 2AFC paradigm is employed the model is constrained to the minimum of 50% correct answers. The delay thresholds are at 65 ms and at 98 ms for the mouse triggered, and for the voice triggered condition respectively.*

***Table 1**: Summary of the results of the threshold determinations in two modalities.*

	VoiVis	**MouVis**
Thresholds obtained with best-PEST (± 95% confidence level)	115 (± 23) ms	78 (± 14) ms
Thresholds obtained from fitted logistic model	98 ms	65 ms
Slopes of the standardized psychometric functions (threshold=0.5)	0.56	0.57
Influence of age (p<0.05)	no	no
Influence of gender (p<0.05)	no	no
Threshold difference due to modality (p=0.01)	VoiVis > MouVis	

As listed in table 1, the threshold of the interaction delay is 78 (±14) ms in the mouse-visual situation, and 115 (±23) ms in voice-visual situation (numbers in brackets stand for the 95% confidence levels). A two-sided, paired t-test shows that the mouse-visual interaction threshold is significantly lower (p=0.01) than the voice-visual interaction threshold. Age and gender do not exert significant influence on the delay thresholds. The slope values of both psychometric functions are essentially the same, when standardised to threshold units.

3. Discussion and Conclusions

Voice-visual interaction delay is less likely to be detected than click-visual interaction delay. This makes sense, since the voice trigger is less distinct compared to the mouse trigger. The chosen adaptive method *best-PEST* has proven suitable for such kind of experiments, because the obtained threshold values differ not significantly from the ones obtained with the model fitting procedure. Nevertheless *best-PEST* exhibits a tendency to overestimate the "true" thresholds, at least with the applied number of trials.

The suggested just noticeable interaction delays may serve as a decision support for network planners and service providers, in such a way that below these values users will definitively not benefit from optimisation of the network referring to round-trip delays. The obtained results are considered as fundamentals for the more complex situation where interaction takes place between humans performing realistic tasks. With realistic conditions contextual as well as psychological factors will play an important role and cannot be controlled so much. Such experiments are subject of our ongoing research.

4. References

1. Bouch, A., Bhatti, N., & Kuchinsky, A. J. (2000). *Quality is in the eye of the beholder: Meeting users' requirements for Internet Quality of Service.* CHI'2000, Hague.
2. Fluckiger, F. (1995). *Understanding networked multimedia: applications and technology*. London: Prentice Hall.
3. Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics, 28*(4), 377-379.
4. Zuberbühler, H. J., Ruegg, S., Krueger, H., & Kündig, A. (2002). *Intermedia Synchronisation in Network Design: Using an Adaptive Psychophysical Method to Specify the Perceivable Audio-Visual Delay.* WWDU 2002 - Work With Display Units: World Wide Work, Berchtesgaden.