Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 66 (2015) 154–175

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Assessment of Cognitive Load, Speech Communication Quality and Quality of Experience for spatial and non-spatial audio conferencing calls

Janto Skowronek *, Alexander Raake

Technische Universität Berlin, Telekom Innovation Laboratories, Assessment of IP-based Applications, Germany

Received 15 April 2014; received in revised form 14 September 2014; accepted 14 October 2014
Available online 23 October 2014

Abstract

The operational characteristics and thus the quality of audio conferencing systems are affected by scalability issues in terms of the number of participants and communication devices, which can change between or even during calls. Towards the development of methods that can handle such scalability aspects, this work investigates the effect of changes in the communicative situation (i.e. Communication Complexity) and the system properties (i.e. Technical System Capability) on the user's quality perception. In a listening-only test, subjects evaluated for a set of artificial telephone conferences their impression of the perceived system performance (i.e. Speech Communication Quality), the perceived effort to follow the conversation (i.e. Cognitive Load), and the overall experience as such (i.e. Quality of Experience). The results showed that not only a technically more advanced system, but also a less complex communicative situation can improve quality perception. Consequently, the assessment of audio conferencing systems needs to incorporate the special communicative situation of such multiparty settings: the methods should avoid confusions between Speech Communication Quality and Quality of Experience; perceptual studies should control for or report on the Communication Complexity to facilitate comparability of studies; and instrumental approaches should incorporate an estimation of the Communication Complexity to improve performance.

Keywords: Multiparty; Conferencing; Quality; Cognitive Load; Scalability

1. Introduction

Scalability of speech technology systems becomes an increasingly important issue. As the term scalable means "capable of being easily expanded or upgraded on demand" (Merriam-Webster, 2014), scalability of a system means to be able to cope with varying quantities of system variables or boundary conditions. Thus, it can refer to the number of users involved, the number of signal channels at system input, transmission path or output, or the number of technical characteristics of a system.

Telephone conference systems are affected by scalability, as they provide a telecommunication platform for multiple users, whereas the number of users and types of communication devices can change between calls or even during a call. Ideally, conferencing systems have to provide sufficient quality irrespective of the number of users and devices, and enable an efficient communication. Telecommunication service operators strive for a sufficient quality by applying quality assessment methods or quality prediction models that enable them to properly plan, monitor and control their systems. However, most of the standardized perceptual assessment methods (e.g. ITU-T Rec. P.805 (2007)) and prediction models (e.g. ITU-T Rec. G.107 (2011)) have been developed for one-to-one conversations;

* Corresponding author.

http://dx.doi.org/10.1016/j.specom.2014.10.003
0167-6393/

Table 1
Overview of the target and manipulation variables and definition of the resulting six hypotheses H1–H6. Each hypothesis can be phrased as "Scaling [Manipulation Variable] has an impact on [Target Variable]".

Target variables	Manipulation variables	
	Communication Complexity (i.e. number of interlocutors *#IL*)	*Technical System Capability* (i.e. sound reproduction method *SndRepr*)
Cognitive Load	Hypothesis H1	Hypothesis H4
Speech Communication Quality	Hypothesis H3	Hypothesis H2
Quality of Experience	Hypothesis H5	Hypothesis H6

their performance for multiparty telephone conferences has not been investigated yet. There is also a first standard for the perceptual assessment of multiparty systems available (ITU-T Rec. P.1301, 2012). This standard document provides a first set of detailed information on a proper conduction of multiparty tests; at the same time the document can often give only general recommendations, e.g. when testing spatial audio conferencing systems. For that reason, more knowledge on this topic is still needed, and multiparty test methods need to be elaborated in more detail and further optimized.

Towards the improvement of such multiparty assessment approaches, the objective of this work is to investigate the effect of scalability issues on the user's quality perception. More specifically, this work will look at scalability issues from both a technical and a communicative perspective: it will investigate the effect of differences in the technical characteristics of the system, expressed as *Technical System Capability*, and the effect of differences in the communicative situation, expressed as *Communication Complexity*. Furthermore, this work will investigate the effect of the mentioned scalability issues on three different aspects of quality perception: the perceived system performance, expressed as *Speech Communication Quality*; the perceived effort to follow the group communication, expressed as *Cognitive Load*, and the overall experience as such, expressed as *Quality of Experience*.

1.1. Communication Complexity

The term *Communication Complexity* refers to the structure of the conversation in terms of who is contributing to the discussion at which point in time. Aspects that differentiate conversational structures are the number of speaker changes, interruptions, monologues, longer speech pauses, etc. How these aspects differ between multiparty conferences depends on how the interlocutors contribute to the conversation. It is assumed in this work, that an interlocutor's contribution to the conversation depends on how he or she is able to perform four mental tasks: (1) understanding the speech signals from the others, (2) identifying speakers and their roles, (3) processing the information shared during the conference and extracting its implications, and (4) formulating adequate responses. Note that this is a simplified model of the cognitive processes inside an interlocutor and should not be interpreted as a comprehensive cognitive model of human-to-human communication. Nevertheless, we can reasonably assume that those four mental tasks do take place, due to the following reasoning.

It is a straight-forward observation that participants perform the first task, and that its difficulty is determined by the speech transmission chain from mouth to ear (e.g. signal quality), the speaking behavior of the interlocutors (e.g. pronunciation), and language aspects (e.g. language fluency). The observation that participants also perform the other three tasks is based on knowledge from the literature on computer mediated group communication and remote collaborative working. For example, Olson and Olson (2000) discusses the importance for the participants to create a common ground, by adapting to what they perceive from the other interlocutors. Other studies (e.g. Sanford et al., 2004; Daly-Jones et al., 1998; Masoodian et al., 1995) also stress the importance of creating such common ground in computer-mediated communication by referring to the work of Clark and Brennan (1991). Similarly, Fussell and Benimoff (1995) discuss that interlocutors "... strive for a shared understanding of the situation, the task, and of one another's background knowledge, expectations, beliefs, attitudes, and the like. They also construct a body of shared knowledge and understanding (common ground) which they can draw upon in their subsequent communications ...".

In terms of scalability, a number of aspects can change *Communication Complexity* by increasing or decreasing the difficulty of accomplishing the four mental tasks. A first aspect is the number of interlocutors. As participants will attempt to understand who the other interlocutors are, and try to adapt their responses, the number of interlocutors will influence the difficulty to identify speakers and formulate responses.

A second aspect is the amount of information shared during a telephone conference. Obviously the more information is shared, the higher is the effort to perform the task of information processing.

A third aspect is the conversational structure in terms of who is contributing to a conversation at which point in time. An increasing complexity in the conversational structure will increase the difficulty of the tasks of speaker identification and information processing, as the order in which people contribute and the order in which information is shared becomes less predictable.