Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 63–64 (2014) 1–14

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

On interaction behaviour in telephone conversations under transmission delay

Katrin Schoenenberg [a,*], Alexander Raake [a], Sebastian Egger [b], Raimund Schatz [b]

[a] Technical University Berlin, T-Labs, TU-Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
[b] Telecommunications Research Center Vienna (FTW), Donau-City-Straße 1, 1220 Vienna, Austria

Received 12 July 2012; received in revised form 9 April 2014; accepted 14 April 2014
Available online 20 April 2014

Abstract

This work analyses the interaction behaviour of two interlocutors communicating over telephone connections affected by echo-free delay, for conversation tasks yielding different speed and structure. Based on a series of conversation tests, it is shown that transmission delay in a telephone circuit does not only result in a longer time until information is exchanged between the interlocutors, but also alters various characteristics of the conversational course. It was observed that with increasing transmission delay, the realities perceived by the interlocutors increasingly diverge. As a measure of utterance pace, a new conversation surface structure metric, the so-called utterance rhythm (URY), is introduced. Using surface-structure analysis of conversations from different conversation tests, it is shown that peoples' utterance rhythm stays rather constant in close-to-natural conversations, but is considerably affected for scenarios requiring fast interaction and a clear answering structure. At the same time, the quality of the connection is perceived less critically in close-to-natural than in tasks requiring fast interaction, that is, interactive tasks leading to a delay-dependant utterance rhythm. Hence, the conclusion can be drawn that the degree of necessary adaption of the utterance rhythm to a certain delay condition co-determines the extent to which transmission delay impacts the perceived integral quality of a call.

Keywords: VoIP; Delay; Interactivity; Conversation analysis; Interaction rhythm; Conversational quality

1. Introduction

The majority of today's telecommunication service providers are increasingly switching to use IP-based transmission as the basis of their telephony services. This change yields a number of advantages, including the ability to implement different services on the same platform, as well as more flexibly using speech codecs, and thus providing a wider speech bandwidth to increase quality. However, with the new technical possibilities new challenges arise, too. A typical problem for IP-based networks is the occurrence of transmission delay. This issue has been known in the context of satellite transmission for many years, even though here transmission delay usually occurred in combination with echo.

In telephone transmission such as Voice over Internet Protocol (VoIP), transmission delays arise at different points in the transmission chain (Thomsen and Jani, 2000; Lakaniemi et al., 2001; Raake, 2006). First, speech encoding and packetization of the recorded speech signal into VoIP packets require buffering of incoming speech samples which causes delay. Then there is, of course, the transmission delay of packets through the network. Third, reaching the receiver side, packets need to be queued in a jitter-buffer before being unpacked, decoded and played out, to ensure

* Corresponding author.
E-mail addresses: katrin.schoenenberg@telekom.de (K. Schoenenberg), alexander.raake@telekom.de (A. Raake), egger@ftw.at (S. Egger), raimund.schatz@ftw.at (R. Schatz).

http://dx.doi.org/10.1016/j.specom.2014.04.005
0167-6393/

a coherent speech signal for the receiver. Due to different routes that packets take through the network, packets may be missing or arrive later. The jitter-buffer tries to re-collect all packets in the sent order. After a predefined time, the not yet arrived packets are usually dropped. In that case, packet-loss occurs. Some jitter-buffers are adaptive in their buffering time. Finally, the packets need to be unpacked and decoded before actual play out. For network planning purposes, it is interesting to predict what effect certain defined delay times have to weight different, for instance, coding methods or buffering times and to most efficiently plan a high quality telephone systems. For this reason, this article will focus on fixed transmission delay times ranging from low to very high values.

As a complement to tests with human subjects, which are time and resource consuming algorithms are being developed to automatically predict speech quality based on transmitted signals or planning parameters (Raake, 2006; Möller et al., 2011). To collect the data required for establishing such prediction models, conversation tests are usually conducted to examine different delay conditions in a controlled manner. Since the delay impairment can only be assessed in conversation-type tests, the International Telecommunication Union (ITU-T) recommends different types of conversation tests and respective conversation scenarios to evaluate the impact of delay on quality (ITU-T, Rec. P.805, 2007). In such tests, people usually accomplish certain predefined tasks under particular transmission conditions. After each task, the test participants are asked to rate the perceived integral quality of the connection on a rating scale, typically the five-point Absolute Category Rating (ACR) scale (ITU-T, Rec. P.800, 1996). The so gathered information is then used to develop or optimise objective quality prediction models such as the E-Model (ITU-T, Rec. G.107, 2009; Möller, 2000).

Scientists in the past and present believe that delay has a critical impact on the quality of a call. However, only one early study (Kitawaki and Itoh, 1991), in which participants were very sensitive to the impairment due to long training and accomplished tasks requiring fast interaction, found people reflecting the degradation in their quality judgement to a larger extent.

Besides assessing quality judgements, various authors showed that the conversation structure changes (Krauss and Bricker, 1967; Brady, 1968, 1971; Hammer et al., 2004; Egger et al., 2010) with increasing delay, which supports the assumption that delay has an impact on communication. For this kind of analysis, conversational states are assigned to the speech on–off patterns derived from recorded conversations according to the commonly used state model (Fig. 1) (Brady, 1965, 1968; Hammer et al., 2004; Hoeldtke and Raake, 2011). Parameters describing the conversation surface structure can then be extracted from the state structure.

This paper aims to resolve how conversation behaviour changes under transmission delay and a perceived degraded quality are related. In particular, we will show

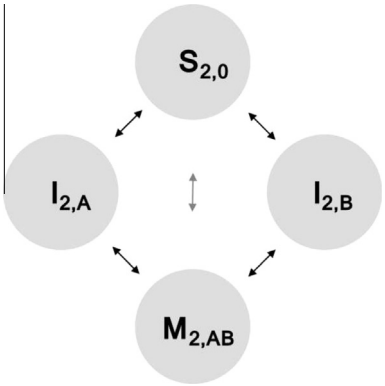

Fig. 1. Possible states (sets and subsets) for a two-party conversation situation; $S_{2,0}$: silence, $I_{2,A}$: individual/single talk of person A, $I_{2,B}$: individual/single talk of person B, $M_{2,AB}$: multi/double talk of person A and B; State transitions indicated by arrows.

under which conversational conditions the quality perception is affected by delay and when this effect is less severe.

In the following, results reported in the literature will be discussed that have studied how the communication style differs when delay is induced versus when no delay is present (Section 2). The approaches considered most beneficial for describing structural changes of interaction behaviour in the context of delay will then be extend and completed. Based on this, the paper presents a series of conversation tests conducted by the authors (Section 3), and uses the results to show how the interaction behaviour changes when conversations are carried out across lines with delay (Section 4). Further, the relationship of the most discriminant interaction parameters and the perceived integral quality ratings collected in the tests is described (Section 4.3). Finally, in Sections 5 and 6, the findings will be discussed and conclusions will be drawn on the questions of which differences can be observed in the interaction structure when delay is present, and under which circumstances delay actually influences the perceived quality.

2. Theoretical background

A number of different approaches have been described in the literature to examine the structure of conversations affected by transmission delay. Table 1 summarizes all conversational parameters considered relevant for this paper, and to be discussed in the following.

2.1. Duration of a call

One of the earliest and most simple parameters used to describe the impact of transmission delay on the call is the duration of a call (*DUR*). Riesz and Klemmer (1963) investigated delay with and without echo and echo suppressors in naturally occurring telephone conversations of laboratory administrative staff. After each call the staff was asked to rate their satisfaction with the call conditions. The results showed a much higher probability for long calls to