The effect of gender and mood change on the perceived integral quality, besides technical conditions in teleconferencing

Katrin Schoenenberg[1], Alexander Raake[1]

[1] *T-Labs, TU-Berlin, 10587 Berlin, Germany, Email: katrin.schoenenberg@telekom.de*

Introduction

In this work we draw our attention to two factors that are usually not considered in conversational quality evaluation of telecommunication services. On the one hand we will examine whether male participants generally give lower or higher ratings in comparison to female participants when they are asked for their perceived integral quality of a call. In most studies gender is balanced across groups, randomly chosen (as it is recommended in ITU-T Rec P.805 [1]) or kept constant without knowing if the gender actually co-determines the quality judgment. For answering this question, we will analyze three studies dealing with the impact of transmission delay on the perceived integral quality.

The second aspect considers the effect of mood change on the perceived integral quality caused by test conduction. As a test supervisor, one often experiences some participants who enjoy conducting the tests and others who seem rather annoyed and negative when conducting them. From psychological research, it is well known that information congruent with the current mood [2, 3, 4] is processed easier and faster. Following from that, judgments can be distorted toward the mood due to selective processing of congruent information. To our knowledge, it has never been studied before if mood affects the perceived integral quality of participants which is why we will address this question here.

Method

For answering the first question, we looked into three studies who have been conducted to assess the impact of transmission delay on the perceived integral quality. In the first study forty-eight participants (24 pairs) conducted conversation tests rating the perceived integral quality on a five-point absolute category rating scale (ARC) after each conversation. This type of rating is usually referred to as the mean opinion score (MOS; ITU-T P.805 [1]). One third of all groups were male-only, one third female-only and the last third mixed in gender. Three types of scenarios were used, the well known 2-party Short Conversation Test (SCTs; [1, 5])), a random number verification task (RNV; [6, 7]) and a timed version of the random number verification (RNT), where a price was promised to the fastest and most correct pair. Since in the first study three different scenario types were used the factor 'scenario' was included in our analysis here. The delay varied in five steps: 100 ms, 225 ms, 425 ms, 825 ms and 1625 ms. All scenario types were combined with all delay conditions (full design).

In study II participants also conducted SCTs, however, here the delay was only varied in three steps: 0 ms, 400 ms and 1200 ms. Every participants talked once to one of three female fixed interlocutors, each time in a different delay condition. All participants were unfamiliar to their interlocutors. In total forty-four participants, resulting in 44 pairs (18 female-only, 26 mixed gender), rated the perceived integral quality. The ratings of the fixed interlocutors are not considered here.

The third study was focussing on the impact of symmetrical and asymmetrical delay conditions in three-party conversations. For this purpose the delay was varied from 0 ms, to 400 ms, to 1200 ms in a symmetrical (all participants delayed) and an asymmetrical (only one participant delayed) manner. All twenty-one participants were unfamiliar to each other, however, were highly familiar with audio-conferencing. On average they reported that they attended 35 multi-party audio conferences in the last year. The seven groups of three consisted of 3 male-only and 4 mixed groups (1 female and 2 males).

In all studies described the conversations were recorded to extract surface parameter from the conversational courses. Particularly, the speech-act rhythm ($SARY$) was calculated because this feature has been shown to be critically related to the perceived integral quality [8]. In this context, we aimed to compare the $SARY$ of males versus females.

To evaluate the impact of mood change on the perceived quality we asked participants to rate their current mood on a seven point absolute category rating scale from ideal to very bad prior to the experiment and after finishing all conversations (study I). The differences of the ratings were calculated, finding either an increase, no change or a decrease in rated mood.

Results

Regarding the effect of gender, two out of three studies showed a significant effect (ANOVAs; study I: $F(1, 652) = 14.98$, $p = 0.000$, study III: $F(1, 115) = 11.64$, $p = 0.001$). Overall the perceived integral quality was rated lower by men compared to women.

For the second study the effect was not significant. However, as depicted in Table 1, the ratings of males were based on mixed groups only. We can see that males, in all studies, tended to rate better if they interacted with females.

All three ANOVAs regarding the effect of gender on the speech-act rhythm $SARY$ turned out to be significant

Table 1: Mean MOS and 95 % CIs; Order of symbols reflects who talked to whom, e.g., ♂ & ♀: a male talked to a female.

Study	Combi.	MOS	Gender	MOS	95 % CI
I	♂ & ♂	2.95	♂	2.97	0.14
	♂ & ♀	3.01			
	♀ & ♂	3.18	♀	3.36	0.15
	♀ & ♀	3.46			
II	♂ & ♀	3.61	♂	3.61	0.20
	♀ & ♀	3.28	♀	3.28	0.32
III	♂ & ♂, ♂	4.10	♂	4.13	0.16
	♂ & ♂, ♀	4.17			
	♀ & ♂, ♂	4.86	♀	4.68	0.21

(study I: $F(1, 655) = 5.79$, $p = 0.016$, study II: $F(1, 130) = 9.97$, $p = 0.002$, study III: $F(1, 133) = 5.97$, $p = 0.016$). In all cases, the men were acting significantly faster than the women on average. For study I, there was also a significant effect for scenario ($F(2, 655) = 190.88$, $p = 0.000$), with the RNV and the RNT scenarios having a faster *SARY* than the SCTs (confirmed by Bonferroni post-hoc tests).

Furthermore, an ANOVA revealed a significant effect for mood change in study I ($F(2,576) = 7.01$, $p = 0.001$). The main effect for scenario and the interaction were not significant. As shown in Fig. 1, people who felt better after the experiment gave an overall higher rating than people who felt worse (confirmed by Bonferroni post-hoc tests). Nevertheless, it looks as if the mood mainly affected the ratings of the number verification tasks. People in general only decreased or increased their rated mood by one.

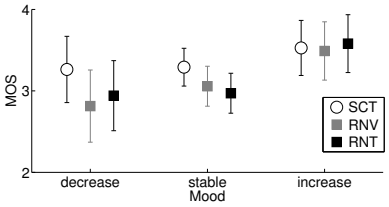

Figure 1: Mean MOS ratings for different mood changes

Conclusion

From the above analysis we conclude that gender matters when conduction conversation test assessing the perceived integral quality for different transmission delay conditions. Males generally rate more critical when they are in a group with only males or at least one other male person. If they talk to one other female only the situation is not as clear, men, in this case, tend to rate higher. This effect may be explained with the effect shown for the speech act-rhythm. Men, on average, begin a new utterance significantly faster than females. Under delay and with an other male as an interlocutor who circles similarly fast this behaviour is likely to lead to, for example, many unintended interruption [7]. In consequence, men may experience the causing technical degradation faster or may develop a generally more negative attitude because the conversations are more exhausting. If the interlocutor of a male is female less confusions [9] may take place under transmission delay because her speech-act rhythm is usually slower. As a result, the quality is not rated as poor by men as in the male-only case.

This work also showed that peoples' quality ratings are affected by their mood increase or decrease caused by the experiment. The effect, however, was mainly due to the game task (RNT and the similar RNV task) where participants could win a price. When delay was present, it was hard to accomplish this kind of task fast enough to feel as being in a winning position. Some participants liked the challenge and rated the quality higher, probably because they were distracted by laughing and enjoying the confusions in the conversations. Other seemed to be frustrated and rated the quality accordingly low. The outcome is in line with our expectations and the findings on the influence of mood in other contexts.

References

[1] ITU-T Rec.P-805: Subjective evaluation of conversational quality. International Telecommunication Union, Geneva, 2007.

[2] Bower, G.H.: Mood and Memory. American Psychologist (1981), 129-148.

[3] Isen A.M. and Daubman, K. A.: The influence of affect on categorization. J. of Pers. and Soc. Psych. (1984) 1206-1217.

[4] Johnson E and Tversky A.: Affect, generalization, and the perception of risk. J. of Pers. Soc. Psych. (1983), 20-31.

[5] Möller, S.: Assessment and Prediction of Speech Quality in Telecommunications. Kluwer Academic Publishers, USA: Boston, 2000.

[6] Kitawaki, N. and Itoh, K.: Pure delay effects on speech quality in telecommunications. IEEE J. on Select. Areas in Commu.(1991), 586-593.

[7] Egger, S., Schatz, R. and Scherer, S.: It takes two to tango - Assessing the impact of delay on conversational interactivity on perceived speech quality. In: Proc of the 11th Annual Conf. of the Int. Speech Commu. Assoc. (2010), 1321-1324.

[8] Raake, A., Schoenenberg, K., Egger, S., Wälterann, M., Schatz, R. and Möller, S.: Sind Sie noch da? - Vorhersage des Einflusses verzögerter Telefonverbindungen auf Konversationsfluss und wahrgenommene Sprachqualität. Fortschritte der Akustik - DAGA (2012).

[9] Brady, P.T.: Effects of transmission delay on conversational behavior on echo-free telephone circuits. Bell System Technical J. (1971), 115-134.