

Elements of Interactivity in Telephone Conversations

Florian Hammer¹, Peter Reichl¹, Alexander Raake²

¹ Telecommunications Research Center Vienna (ftw.), Austria

² Institute of Communication Acoustics (IKA), Ruhr-University Bochum, Germany

{hammer|reichl}@ftw.at, alexander.raake@ruhr-uni-bochum.de

Abstract

The term “interactivity” has been defined in numerous ways in the context of communications, but a definition of interactivity as an instrumentally measurable parameter of conversations is still missing. In this paper, we approach this issue by applying a parametric analysis to telephone conversations recorded during speech quality tests. To this end, we extract the basic conversational parameters like speech activity, mutual silence and double talk as well as a set of conversation events like speaker alternation rate and interruption rate. Comparing two types of scenarios for conversational speech quality assessment and exploring four different situations with regard to transmission delay, we aim at understanding the interdependencies between the conversational parameters which are basic for studying interactivity. This is intended to, ultimately, lead to an instrumental metric distilled from the relevant parameters.

1. Introduction and Motivation

In today’s networked multimedia world, the concept of “interactivity” has many different faces. In this work, we focus on interactivity as an instrumentally measurable characteristic quantifying the dynamic quality of telephone conversations. We can identify two major issues that would benefit from a sharp definition of interactivity with regard to conversational speech quality assessment of telephone connections. First, with such a definition at hand, the interactivity of the tasks that test persons need to fulfill during the conversational tests can be compared, which is especially interesting for tests involving transmission delay. Second, a metric for interactivity would allow the connection impairments which influence the temporal behavior (delay and its variation) to be put in relation to the subjective rating that is given by the subjects on the quality of the connection. The latter issue is a topic in the ongoing discussion of how the E-model [1], the telephone network planning model standardized by the ITU-T (International Telecommunication Union), should deal with impairment due to delay, since high latency values do not seem to degrade the user opinion as much as expected on echo-less connections [2].

This paper describes the first cycle of an iterative approach, with the final aim of defining “interactivity”, and of developing a metric quantifying the impact a particular transmission system exerts on it. The present iteration consists in approaching (but not yet defining) such an interactivity metric by applying Parametric Conversation Analysis (PCA), with a three-fold motivation: 1) Evaluate the PCA method using recorded conversations obtained using two different types of conversation tasks; 2) evaluate the extracted interaction parameters in the light of the *interactivity* measure aimed at; 3) evaluate whether the conversation task types differ in the interaction parameters, potentially predestinating one of the two for the subjective assess-

ment of speech conversation quality under transmission delay. In future iterations, the measurement set-up as well as the interactivity metric itself will be refined.

The paper is structured as follows. First, we will briefly review related work. In section 3, we will describe the parameters we use to formally characterize a conversation. The test method is described in section 4. Section 5 presents and discusses results from the analyses of conversation recordings. Finally, we conclude the paper in section 6.

2. Related Work

2.1. Conversational Parameters

Brady [3] has been among the first ones to provide a detailed analysis of conversational parameters, in this case extracted from 16 conversations. Brady has defined “conversational events” like talk-spurt, mutual silence, double talk, and interruptions. For collecting the speech material, the test persons (close friends) were asked to talk about anything they wished. Brady extracted the parameters automatically using a threshold-based detection algorithm that excluded talk-spurts smaller than 15 ms and filled pauses smaller than 200 ms.

A method for *generating* artificial conversational speech patterns is described in ITU-T Rec. P.59 [4]. This standard presents a four state model of conversations and the corresponding temporal parameters averaging values obtained for English, Italian, and Japanese. We will present these values in the results section for comparison with our own results.

2.2. The Concept of Turn-taking

The parameters described in the previous section are related to talk-spurts. At the next level of analysis, we can distinguish between either speaker A or B (cf. Figure 1) having the conversation floor, which leads us to the concept of turn-taking (cf. Sacks et al. [5]) commonly used in “traditional” Conversation Analysis (CA). In CA, semi-verbal utterances like “mhm”, “uh-huh”, or laughter are regarded to as back-channeling events which establish rapport between talker and listener and encourage the talker to continue. In CA, back-channeling does not necessarily constitute turns. However, we have incorporated these utterances into our investigations as talk-spurts since a semantic analysis of the speech signals is out of the scope of this work. Furthermore, we assume these events to increase the liveliness of the conversation.

3. Parametric Conversation Analysis

In this section, we present the concept of *Parametric Conversation Analysis* (PCA). The term “parametric” refers to the fact that this concept is based on parameters that can be extracted in-

strumentally from conversation recordings. Moreover, this term allows the distinction from “traditional” CA which mainly focuses on the semantic aspect of conversations. PCA deals with the parameters of a 4-state conversation model and a set of conversational events which are expected to correlate with interactivity.

3.1. Conversation Model

According to the model used here, a two-way conversation can be divided into four different states (cf. [4]), as illustrated in Figure 1.

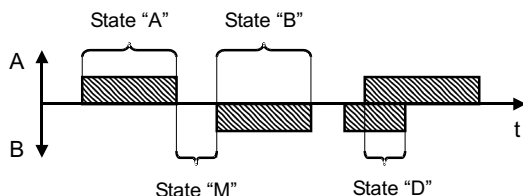


Figure 1: Illustration of the conversational states.

States A and B denote that either person A or person B is talking and the other person does not speak. State M (mutual silence) reflects the situation that both persons are silent, and state D (double talk) represents the case that both persons talk simultaneously.

This behavior can be modeled as a 4-state Markov process (cf. [4]) which is depicted in Figure 2. Note that the transitions between A and B and the transitions between M and D are negligible due to their rare occurrence.

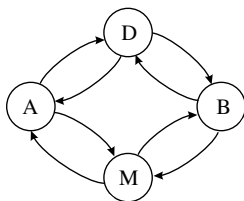


Figure 2: Model of the basic conversational parameters.

The model is described by the transition probabilities between the four states. Furthermore, the sojourn times t_A, t_B, t_M, t_D represent the average times that the process stays in the corresponding state, whereas p_A, p_B, p_M and p_D refer to the overall probabilities to be in one of the states A, B, M and D , respectively.

3.2. Conversational Events

More details on the characteristics of a conversation are given by the conversational events (cf. [3]). The events analyzed in this paper are explained in the following.

The *speaker alternation rate* (SAR) represents the number of alternations between the talkers per minute. Alternations include either state sequences in which the talk spurts are separated by mutual silence (A-M-B and B-M-A), or sequences like A-D-B and B-D-A in which one talker is interrupted by the other. The speaker alternation rate can be compared to a turn rate in CA.

We consider an interruption as a special case of alternation and define the *Interruption Rate* (IR) as the number of interruptions per minute. Note that at higher delay values, unintended

interruptions may occur if the talkers do not adapt their conversation discipline to the connection.

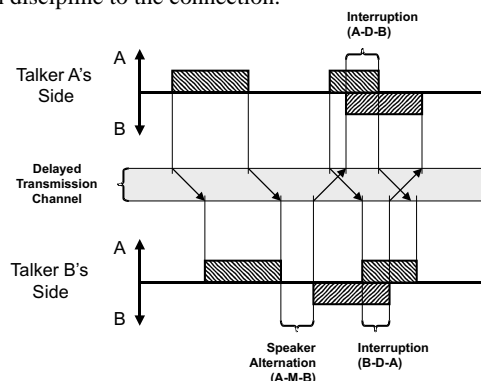


Figure 3: Illustration of Conversational Events.

Figure 3 illustrates the speaker alternation and interruption events. Note that the interruptions occur as a result of the transmission delay. Talker B responds to talker A 's first spurt and unintentionally interrupts talker A who might have expected an earlier response of talker B . Thus, the amount of double talk is increased. Moreover, the figure illustrates the increase of mutual silence due to delay. The speaker alternation duration talker B 's side increases by two times the one-way delay compared to a zero-delay connection. In our particular case, this duration is reduced by the occurrence of the second talk-spurt of talker A .

An analysis of the the mean durations of speaker alternation and interruptions and the amount of non-interruptive double-talk is not performed in this paper but is considered for future work.

4. Test Method

4.1. Conversation Scenarios

As a part of a test method to determine the conversational parameters it is necessary to involve the conversation partners in an appropriate conversation task using predefined *conversation test scenarios*. Different types of conversation scenarios have been described in the literature [6, 7]. The tasks range from interactive games (e.g. TNO test tasks, [6]) over identifying differences between two versions of pictures (e.g., described in [7, pp. 75-77]) to the rapid exchange of random numbers.

The main shortcomings of many of the above scenarios is the lack of naturalness for telephony contexts—reading random numbers or playing a game do not represent an everyday telephone usage. Therefore, the so-called Short Conversation Tests (SCTs) have been developed at IKA [7]. They represent real-life telephone scenarios like ordering a pizza or reserving a plane ticket, leading to comparable and balanced conversations of a short duration of 2–3 minutes (as opposed to 8–10 minutes in case of most scenarios mentioned above). The SCT scenarios are now commonly used in conversation tests carried out in the framework of the speech quality related activities within standardization bodies like the ITU-T.

In order to yield a measurable effect of transmission delay on the conversation parameters (cf. section 3) or on user opinion, the applied conversation tasks need to be sufficiently interactive. When using the SCT scenarios, even one-way transmission delays of up to 1000 ms only had a minor effect on user opinion (e.g. [7, 2]). Therefore, in addition to the existing set of SCT scenarios, interactive Short Conversation Test scenarios (iSCT scenarios) were developed at IKA for the present study,

Table 1: State probabilities [%] for the SCT and iSCT scenarios and ITU-T Rec. P.59 [4].

Scenario	State Probabilities [%]			
	A	B	M	D
SCT	34.3 (7.1)	34.8 (3.6)	27.3 (8.3)	3.6 (2.2)
iSCT	37.83 (6.01)	38.72 (6.99)	18.58 (5.98)	4.86 (1.76)
P.59	35.24	35.24	22.48	6.59

which were expected to yield more interactive conversations. They were chosen as a combination of natural, “telephone typical” situations with simple interactive tasks.¹

The iSCT tasks consist in the rapid exchange of numerical or lexical data motivated by the scenario, such as exchanging room numbers and email-addresses of new employees of a large company. In order to prevent the subjects from too quickly applying a “walkie-talkie”-like strategy (only one subject speaks at a time), not all information required by one of the conversation partners is made available. In this way, additional turn-taking is provoked.

In this study, the PCA method is applied to conversations obtained using both the SCT and the iSCT scenarios. To increase interactivity during the tests involving the iSCT scenarios, the conversation discipline was lowered by only selecting pairs of subjects who knew each other well, and by instructing them to call each other by their first names. To further increase interactivity, the subjects were instructed to perform the tasks as quickly as possible.

4.2. Conversation Recordings and Talk-spurt Extraction

The test set-up for conversational speech quality measurement at IKA is described in [7]. The set-up consists of two telephone handsets connected to an ISDN simulation system including components for the emulation of voice-over-IP transmission and the injection of transmission delay. We have recorded the microphone signals of both talkers for further investigations. Each conversation recording has been cut as to start with the called person picking up the hand-set and end with the second person hanging up. In order to be able to calculate the conversational parameters of interest, the conversation talk-spurts have been marked manually in CoolEdit following Brady’s rule including pauses smaller than 200ms into the talk-spurt. These cues have been extracted and transferred to MatLab for further processing.

5. Results

In this section, we present the results of the PCA of the recorded conversations, comparing the SCT and iSCT scenarios, and illustrating the effect of transmission delay on the parameters.

5.1. Comparison of Scenarios

We compare the SCT and iSCT scenarios (10 conversations per scenario) and ITU-T Rec. P.59 [4] by exploring the corresponding state probabilities and sojourn times of the conversational

¹The naturalness of the scenarios was verified by interviewing the subjects after the entire conversation test (“Conversations perceived as natural?, Yes/No”). In case of the iSCTs, a similar percentage of “Yes” answers as for the SCTs was obtained (app. 78% vs. 83%)

Table 2: Sojourn Times for the SCT and iSCT scenarios and ITU-T Rec. P.59 [4].

Scenario	Sojourn Times			
	A	B	M	D
SCT	1.45 (0.23)	1.59 (0.43)	0.68 (0.13)	0.35 (0.08)
iSCT	1.44 (0.42)	1.44 (0.40)	0.42 (0.09)	0.33 (0.08)
P.59	0.78	0.78	0.51	0.23

model, as given in Tables 1 and 2, respectively. For the comparison of the SCT and iSCT scenarios, clean PCM a-law encoded connections were analyzed. To increase comparability, one specific SCT scenario was compared to one specific iSCT scenario.

The most significant difference between the SCT and iSCT scenarios is reflected in the parameter values for mutual silence being much smaller for the iSCT scenario. We can observe the same behavior inspecting the sojourn times. A reduction of mutual silence may indicate higher interactivity. In comparison, the probability of double talk shows no significant increase for the iSCT scenario whereas both kinds of scenarios result in significantly less double talk than given in P.59. This may be due to the structure provided by both the SCT and iSCT conversation tasks.

Talk spurts of iSCTs tend to occur more often than those of SCTs but they are about equally long. The sojourn times for mutual silence and double talk are basically the same for both SCTs and iSCTs. The latter fact holds true also for the parameters given by ITU-T Rec. P.59, whereas the mean talk-spurt durations of P.59 are significantly smaller. This difference may result from the fact that our conversations were held in German, and P.59 is based on English, Japanese, and Italian speech material.

Table 3: Mean values and standard deviations of speaker alternation rate (SAR), and the number of interruptive for both SCTs.

Task	SAR	IR
SCT	19.66 (4.62)	4.28 (2.57)
iSCT	26.04 (5.59)	5.91 (1.63)

Table 3 presents a comparison of the speaker alternation rates of the SCT and iSCT scenarios. As expected, the iSCT scenarios result in an increase in speaker alternation rate. In addition, speakers tend to interrupt themselves more often in iSCT scenarios which may also give rise to its increased double talk probabilities.

5.2. The Effect of Delay

The effect of one-way transmission delay on the conversational parameters is studied for ITU-T G.729 encoded connections at four different conditions: 60 ms, 360 ms, 660 ms, or 960 ms.

The state probabilities for the delay conditions are depicted in Figure 4. From this figure, we can make two major observations: First, the probability for mutual silence increases between 360ms and 660ms which may indicate that the test persons notice the delay and adapt their talking behavior. The increase in mutual silence seems to result from the talkers speaking less of-

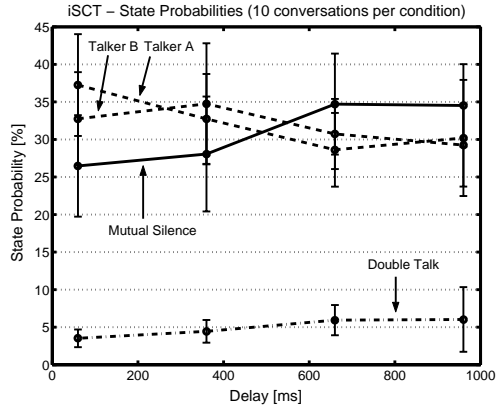


Figure 4: State probabilities at different delay values.

ten resulting in reduced amount of talk-spurts. Second, double talk does not, as expected, increase at very high delay values. The first observation seems to reflect times of transmission at which nobody talks.

Figure 5 depicts the average speaker alternation and interruption rates. While the speaker alternation rate is significantly reduced at low delay, it remains constant as the delay is further increased. The lower speaker alternation rates may be related to the higher amount of mutual silence than observed above. The number of interruptions at a particular talker's end (denoted as "near-end" in the Figure) does not vary significantly with delay. In contrast, the rate of interruptions caused by the other talker at the "far-end" increases. This effect is related to the loss in conversation structure caused by the delay at the far end.

As a final result, we have observed that the mean call durations, 122.35 s, 141.26 s, 154.25 s, and 155.43 s, for delays of 60 ms, 360 ms, 660 ms and 960 ms, respectively, seem to saturate at high delay values.

6. Conclusions

Within the framework of "Parametric Conversation Analysis", we have proposed a method for the analysis of conversational parameters of telephone conversations which allows for the development of a metric for interactivity. We have applied this method as to compare two kinds of short conversational tests for conversational speech quality assessment of which one (iSCT) is assumed to be more interactive. Our results indicate that the iSCTs are indeed more interactive.

As the most important parameters affected by the scenarios, the state probabilities for mutual silence (M) and for double talk (D), the sojourn times for mutual silence t_m and the speaker alternation rate (SAR) can be identified. The main parameters affected by delay are state probabilities p_A , p_B , p_M , and p_D , as well as the alternation and interruption rate. The increase of p_M and p_D reflects the loss of conversation structure. The increase of p_M is partly due to the changes in the reactions of the two interlocutors, and partly due to situations as depicted in Fig. 3.

In summary, the interruption rate at the far end, as well as the state probabilities p_M and p_D seem to qualify as candidates for the interactivity metric, as they prove to be affected by both the scenarios and the delay. The SAR as additional parameter seems to be applicable as a binary indicator for interactivity, switching between values of $SAR \approx 26$ [1/min] for more interactive tasks to $SAR \approx 19$ [1/min] for additional delay ($T_a > 300$ ms) or less interactive tasks.

Future work includes an analysis of the correlation of the

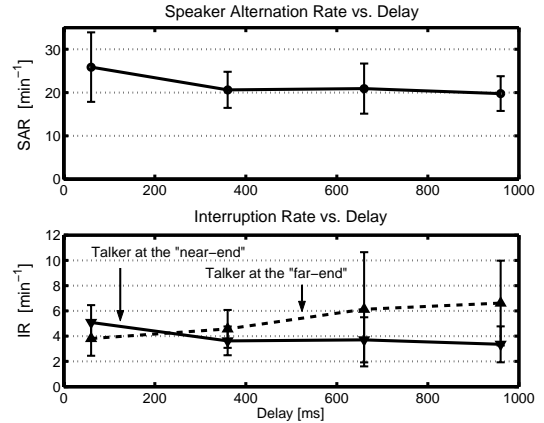


Figure 5: Comparison of the performance of the speaker alternation rate (SAR) and the Interruption rate at different delay values.

conversational parameters and the subjective ratings given for the individual telephone connections. The design and conductance of further conversation experiments will provide detailed information about the perceived interactivity in the context of telephony. Based on such data, we aim to develop an instrumental metric for interactivity.

7. Acknowledgements

This work has been funded under the Austrian government's Kplus Competence Center Program and partly funded by the EC in the framework of the IST project INSPIRE (IST-2001-32746). The authors would like to thank Gernot Kubin for the fruitful discussions and Stefan Schaden for the valuable input.

8. References

- [1] International Telecommunication Union, "The E-model, a computational model for use in transmission planning", ITU-T Rec. G.107, Mar. 2003.
- [2] Raake, A., "Predicting speech quality under random packet loss: Individual impairment and additivity with other network impairments", Acta Acustica, to appear, 2004.
- [3] Brady, P. T., "A statistical analysis of on-off patterns in 16 conversations", Bell Syst. Tech. J., 47(1):73-91, 1968.
- [4] International Telecommunication Union, "Artificial conversational speech," ITU-T Rec. P.59, Mar. 1993.
- [5] Sacks, H., Schegloff, E. A., and Jefferson, G., "A simplest systematics for the organization of turn-taking for conversation", Language, 50(4):696-735, 1974.
- [6] Wijngaarden, S. J. v., et al., "Communicability Testing for Voice Communications", IEEE Workshop on Speech Coding, Ibaraki, Japan, Oct. 2002.
- [7] Möller, S., Assessment and Prediction of Speech Quality in Telecommunications, Boston: Kluwer, 2000.