Hot Discussions and Frosty Dialogues: Towards a Temperature Metric for Conversational Interactivity

Peter Reichl and Florian Hammer

Telecommunications Research Center Vienna (ftw.), Donaucitystr. 1, A-1220 Vienna, Austria
{reichl | hammer}@ftw.at

Abstract. *Despite of its evident importance, measuring interactivity as a central parameter for characterizing conversations has not experienced much attention so far. Starting with an axiomatic approach and using thermodynamic concepts, this paper proposes "conversational temperature" as a new interactivity metric. The resulting scalar parameter is easily to be determined from standard measurements with conversation traces. Several flavors of the method are introduced and investigated for different tasks through a variety of experiments, thus validating the efficiency of our proposal.*

Keywords: Conversational nteractivity, perceptual speech quality, conversation experiments, E-model, socio-physics

1 Introduction

With the imminent introduction of Quality-of-Service (QoS) enabled services in packet-based networks like the Internet, the question of how to evaluate the user's perception of quality has gained rapidly increasing interest both in industry and academia. In this context, the most popular framework comes from ITU-T, where the so-called "E-model" has been standardized as a new concept of impairment factors for predicting the perceptive effects of different types of degradation like delay, jitter or datagram loss, on the overall speech communication quality [5]. Among the conversational factors not yet included, the "interactivity" of a conversation might have important consequences especially concerning delay and jitter.

Providing a satisfying definition of interactivity has turned out to be a non-trivial task, despite the unquestionable existence of a strong common feeling about this concept [3]. It is interesting to observe, though, that in the absence of a formal description we often resort to temperature as an appropriate parameter for capturing the liveliness of conversations, which in this sense virtually may range between "hot" discussions and "frosty" dialogues. Therefore, in this paper we refrain completely from the definition problem and focus instead on the axiomatic foundation of a scalar interactivity metric, the "conversational temperature", to be derived with the help of thermodynamic concepts. To our knowledge, this is a new approach, which admittedly has some parallels in the area of "socio-physics", e.g. Sznaid's consensus model [9] as an application of the Ising model to the problem of opinion formation in closed communities.

The remainder of this paper is structured as follows: Section 2 presents a general state model of conversations and formulates three axioms for an interactivity metric. Section 3 describes norm conversations and defines "conversational temperature" in close analogy to the corresponding thermodynamic approach. Section 4 validates our proposal with a couple of measurement results, before Section 5 presents a slight generalization using Stochastic Petri Nets. Section 6 concludes the paper with some final remarks.

2 An Axiomatic Definition of Interactivity

2.1 Conversation Model

In contrast to the classical study of Brady [2] who distinguishes not less than 10 relevant "conversational events", we follow [4] and use the reduced model of Figure 1 which sketches an arbitrary conversation between two speakers in terms of four states A, B, M ("mutual silence") and D ("double-talk"), depending on whether speaker 1 and/or 2 is active or silent (see Table 1). For each state $I \in \{A, B, M, D\}$, define t_I to be the average sojourn time, i.e. the average time spent in this state.

Table 1: Four States of a Conversation

	# 1 silent	# 1 active
# 2 silent	**M**	**A**
# 2 active	**B**	**D**

Later on we interpret Figure 1 as a 4-state Markov process, where the lack of direct transitions between states *A* and *B* or between state *M* and *D* illustrates the fact that these are quite rare events (e.g. usually each talk spurt is either followed by mutual silence or interrupted by a period of double-talk).

2.2 Axioms of Conversational Interactivity

This paper presents an approach to derive a uniform parameter $\tau = \tau(t_A, t_B, t_M, t_D) \geq 0$ as function of the mean sojourn times allowing a simple but efficient one-dimensional description of conversational interactivity. As mentioned in Section 1, related work does not provide a unified definition of interactivity. Therefore, in the following we chose an axiomatic approach eventually leading to an implicit definition.

Axiom I (Zero Interactivity):
A conversation has zero interactivity ($\tau = 0$) iff $t_I \to \infty$ for one or more states $I \in \{A, B, M, D\}$.

The "Zero Interactivity Axiom" is justified by the observation that the interactivity of a conversation is low (tends to zero) if

- the speakers are speaking for long periods (unidirectional information flow, e.g. during speeches or presentations), or
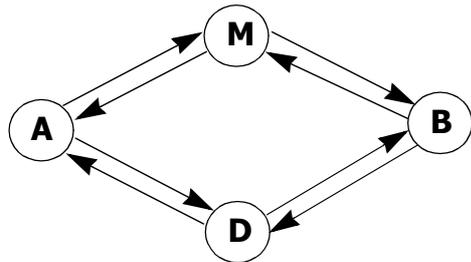

Figure 1: General Conversation Model

- both speakers remain silent for long periods (no information flow), or
- both speakers are speaking at the same time for long periods (lack of information reception, often found in talk shows or TV discussions among politicians).

As a consequence, Axiom I allows us to describe the limiting behaviour of $t_I(\tau)$ for $\tau \to 0$ as follows:

$$\lim_{\tau \to 0} t_A(\tau) = \lim_{\tau \to 0} t_B(\tau) = \lim_{\tau \to 0} t_M(\tau) = \lim_{\tau \to 0} t_D(\tau) = \infty \quad (1)$$

Axiom II (Monotonicity):

$$t_I < t_I' \Leftrightarrow \tau(t_I, t_{-I}) > \tau(t_I', t_{-I}) \text{ for any } I \in \{A, B, M, D\} \quad (2)$$

Also the "Monotonicity Axiom" may be justified empirically: As soon as one of the four states is visited for a shorter mean time, the vividness of the conversation increases, and hence the interactivity metric τ. Note that according to standard game-theoretic notation, for any $I \in \{A, B, M, D\}$, t_{-I} corresponds to the set of mean sojourn times of all states except I.

Axiom III (Normalization):

Average conversations have average interactivity.

The "Normalization Axiom" is a simple requirement for enabling a user-based evaluation of the interactivity metric where it is necessary to offer the test subjects a appropriate (MOS) scale in order to receive meaningful results.

Summarizing the axioms, our search for candidates for $t_I(\tau)$, $I \in \{A, B, M, D\}$, has to focus on positive real functions that are defined for $\tau \in (0;\infty)$, converge to ∞ for small τ and are monotonically decreasing in τ. In the following sections we will argue that indeed

$$t_I(\tau) \sim \exp\left(\frac{\text{const}}{\tau}\right). \quad (3)$$

is a good candidate function. Furthermore, the ultimate form of $t_I(\tau)$ will be derived, and we will demonstrate the validity of our choice through a couple of examples.

3 Markov Model and Thermodynamics

3.1 Parametrization of a Norm Conversation

Conversations viewed as a sequence of sojourn times t_I spent in states $I \in \{A, B, M, D\}$ may cover a huge variety of potential realizations, depending e.g. on language, task, temperament, situation etc. of the speakers. For the time being, we abstract from all these marginal conditions and introduce as "norm conversation" the one described in [4] by specifying average conversation parameters for a mixture of English, Italian and Japanese. These parameters include the steady-state distribution $\bar{\pi}_I$, transition probabilities $\bar{\pi}_{IK}$ and mean sojourn times $\bar{t}_I$ for states $I, K \in \{A, B, M, D\}$. E.g.

$$\bar{t}_D = 0.228, \bar{t}_M = 0.508 \quad (4)$$

are specified directly in [4], whereas the average length of talk-spurts (1.004 sec) and pauses (1.587 sec) is described from the perspective of an isolated speaker, yielding a talk-spurt rate of 38.7% and a pause rate of 61.3%. The double talk rate (over the whole conversation) is specified to be $\bar{\pi}_D = 6.6\%$, the mutual silence rate to be $\bar{\pi}_M = 22.5\%$. Hence, in a norm conversation, $\bar{\pi}_D \cdot 38.7\% = 17.1\%$ of a talk-spurt is spent under double-talk, whereas for 82.9% of the talk-spurt

duration, the opponent is silent, yielding an average sojourn time for state A (for symmetry reasons also for state B) of

$$\bar{t}_A = \bar{t}_B = 0.832. \quad (5)$$

3.2 Definition of "Conversational Temperature"

If we interpret Figure 1 as a regular Continuous Time Markov Chain (CTMC), then t_I, the sojourn time in state I, is exponentially distributed [7] with parameter

$$\lambda_I = 1/t_I . \quad (6)$$

Note that λ_I can be interpreted as total transition rate out of state I [7], and the transition rates λ_{IK} from state I to state K can be calculated for $I, K \in \{A, B, M, D\}$, $I \neq K$, as

$$\lambda_{IK} = \lambda_I \cdot \pi_{IK} = \frac{\pi_{IK}}{t_I} . \quad (7)$$

Now we risk a plucky view into thermodynamics by drawing an analogy between conversations running along a sequence of the four basic states and a particle moving within a container that is subdivided into four "wells" (as known from quantum physics) which are separated from each other by energy barriers ΔE (see Figure 2). Then, we can view the particle as a kind of "state token" residing in one of the wells (states) and jumping over the respective barriers into an adjacent well from time to time. Now assume the token is in state I and tries to jump over the energy barrier. Then, a well-known result from thermodynamics (cf. e.g. [8]) states that the success rate λ_I equals

$$\lambda_I = \nu_I \cdot \exp\left(-\frac{\Delta E}{k\tau}\right) \quad (8)$$

where ν_I is a constant (oscillation frequency = rate of jump trials) characterizing the token behaviour, k is the so-called Boltzmann's constant and τ refers to the absolute temperature of the system. In our analogy, we can normalize $\Delta E = 1$, and using (6) we get

$$t_I = \frac{\exp(1/(k\tau))}{\nu_I} . \quad (9)$$

Next, we follow Axiom III and define the norm conversation to have "room temperature" τ_0 (e.g. in human terms $\tau_0 = 20°$, in a certain sense corresponding to 293 K). This determines the constant ν_I as

$$\nu_I = \frac{\exp(1/k\tau_0)}{\bar{t}_I} \quad (10)$$

Now, we can use (8), (9) and (10) to describe t_I, i.e. the actual mean sojourn time of a conversation in state I, as a function of τ:

$$t_I(\tau) = \frac{\exp(1/(k\tau))}{\nu_I} = \exp\left(\frac{1}{k\tau} - \frac{1}{k\tau_0}\right) \cdot \bar{t}_I. \quad (11)$$
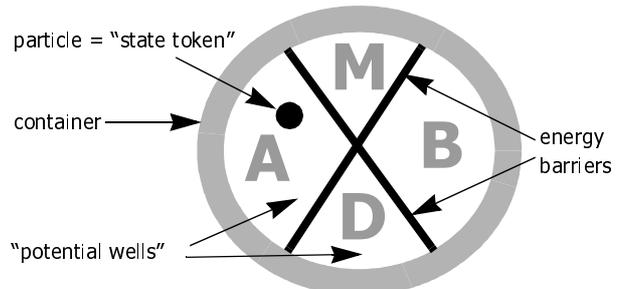

Figure 2: The Quantum Well ("Potential Well") Analogy

Eventually, we have to determine the equivalent κ of Boltzmann's constant. To this end, it seems reasonable to linearize the system around the temperature τ_0 such that a uniform deviation from the norm sojourn times by $(1/\tau_0) \cdot 100\%$ should be reflected by a temperature change of $1°$ C in the opposite direction. Thus, from (11) we get

$$\left(\frac{\tau_0 - 1}{\tau_0}\right)\bar{t}_I = \exp\left(\frac{1}{\kappa(\tau_0 + 1)} - \frac{1}{\kappa\tau_0}\right) \cdot \bar{t}_I \qquad (12)$$

for any I and therefore

$$\frac{1}{\kappa} = -\ln\left(\frac{\tau_0 - 1}{\tau_0}\right) \cdot \tau_0(\tau_0 + 1). \qquad (13)$$

It is worth noting that in the limiting case $\tau \to 0$ where interactivity completely vanishes, (11) complies exactly with our former characterization of this case as given in (1).

But we can use (11) also the other way round, defining the "conversational temperature" τ as a function of the measured mean sojourn times t_I. Then, we can provide an estimate $\hat{\tau}$ of the temperature τ for example through weighted least-square fitting:

$$\hat{\tau} = \text{argmin}\left\{\sum_I \alpha_I\left(\bar{t}_I \exp\left(\frac{1}{\kappa\tau} - \frac{1}{\kappa\tau_0}\right) - t_I\right)^2\right\}. \qquad (14)$$

In the simplest case, we set $\alpha_I = 1$ for $I \in \{A, B, M, D\}$, but we can also use more sophisticated weights, e.g. to put emphasis on the role of double-talk etc.

In any case, (14) assigns each conversation a uniquely defined "temperature" τ, based on (11) and the measured average mean sojourn times t_I. Note that another (potentially more accurate) way to estimate the conversational temperature takes also the transition rates (7) into account and therefore requires additionally the transition probabilities π_{IK} of the embedded Markov chain describing the conversations to be measured. In this case, (8) becomes

$$\lambda_{IK} = \pi_{IK} \cdot \nu_I \cdot \exp\left(-\frac{1}{k\tau}\right), \qquad (15)$$

where (10) and (13) remain valid for ν_I and κ, resp. Thus, the new temperature estimator is

$$\hat{\tau}' = \text{argmin}\left\{\sum_{\substack{I, K \\ I \neq K}} \alpha_{IK}\left(\frac{\pi_{IK}}{\bar{t}_I}\exp\left(\frac{1}{\kappa\tau_0} - \frac{1}{\kappa\tau}\right) - \frac{\pi_{IK}}{t_I}\right)^2\right\} \qquad (16)$$

4 Examples and Results

4.1 Experiment Setup

To validate our approach, we have applied it to a series of conversation experiments performed at the Institute of Communications Acoustics at Ruhr-University Bochum, Germany. The detailed setup is described in [6]. The test persons had to perform two types of tasks, i.e. (1) short conversational tests (SCT) like ordering pizzas or booking flights, and (2) interactive short conversational tests (iSCT), e.g. fast exchange of meteorological data or telephone numbers. Note that the iSCTs have been designed to be significantly more interactive than the SCTs. In this paper, we use conversation reordings from 9 pairs of test persons having had three conversations each.

4.2 Temperature of Artificial Speech

One way to illustrate visually the effect of increasing conversational temperature is depicted in Figure 3. Here, average conversation parameters measured for the above "Pizza" sce-
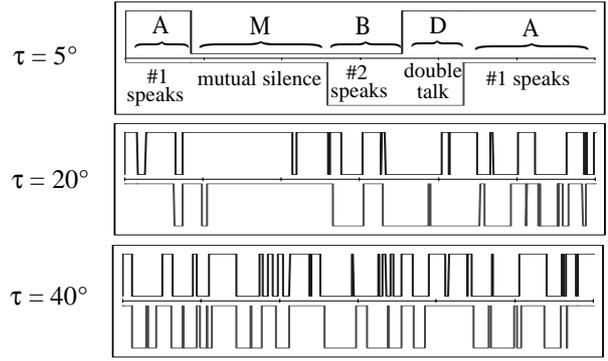

Figure 3: Effect of Temperature for Artificial Speech

nario are used to create artificial conversation patterns of length 60 secs for temperatures of $5°$, $20°$ and $40°$, resp. Speaker 1 activity is depicted above the x-axis, Speaker 2 below, and the x-axis itself corresponds to periods of silence.

4.3 Measurement Results

Having illustrated the effect of temperature on simulated conversation patterns, now we present some results for the experiments described in Section 4.1. Table 2 deals with the average temperature τ according to (14) for two selected scenarios, i.e. the comparison of meteorological data and the process of ordering a menu from a pizza service, compared to $\tau_0 = 20°$ for the norm conversation as described in Section 3.1. Apparently, the overall temperature achieved in our conversation experiments is much lower than τ_0, moreover the less interactive iSCT (pizza) leads also to a noticeable lower temperature compared to SCT (weather)..

Table 2: Conversational Temperature

Task	Average Temperature τ	Standard Deviation
Weather Data	13.3	1.22
Pizza Service	12.8	1.19

Finally, we compare approaches (14) and (16), each of them either with equal weights $\alpha_I = \alpha_{IK} = 1$ (EW) or with weights according to the steady-state distribution $\alpha_I = \alpha_{IK} = \pi_I$ (i.e. π_I as probability that the conversation is in state I (SP). Assuming $\tau_0 = 12.8°$ (cf. Table 2) for the weather scenario as new reference case, Table 3 presents the results for the pizza case. Again, the temperature for the more interactive iSCTs (weather) is significantly higher than for the SCTs (pizza), while the results according to (16) further support this trend corresponding to our a priori assumptions due to the experiment design (see Section 4.1).

Table 3: Relative Weather Temperature (Pizza = 12.8°)

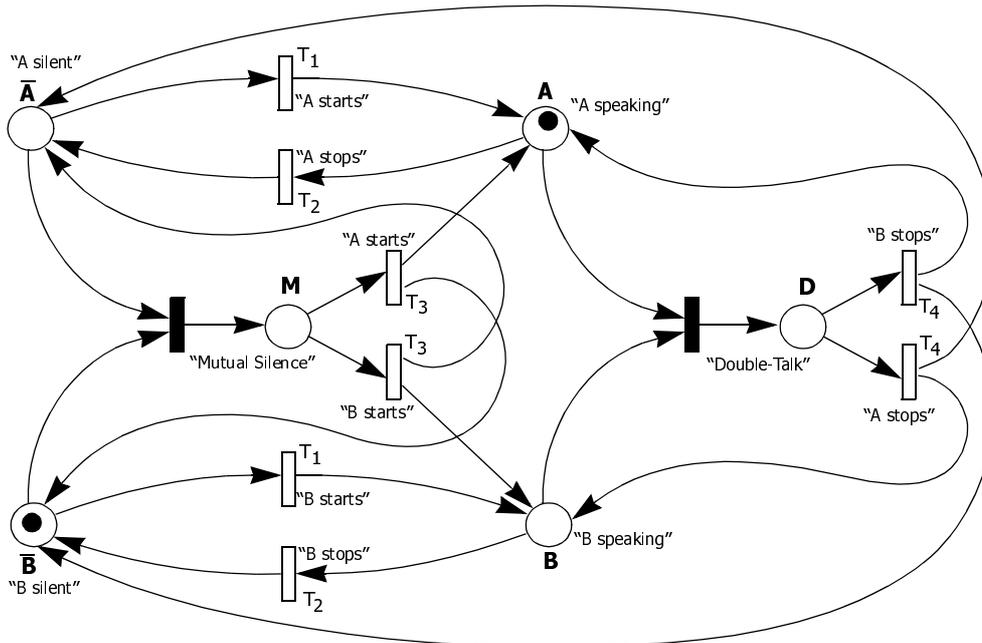| Method	Average Temperature τ	Standard Deviation
(14) + EW	14.6	3.33
(14) + SP	14.0	3.33
(16) + EW	15.6	1.44
(16) + SP	16.9	2.30

Figure 4: Stochastic Petri Net Model for Conversations

5 A Stochastic Petri Net (SPN) Model

In the previous section, we have introduced "conversational temperature" as a uniform parameter influencing the transition rates of a Markov process. One of the drawbacks of this approach concerns the fact that Markov modeling requires the fundamental assumption that all sojourn times are exponentially distributed. In order to extend our approach for more general distributions, Figure 4 describes a conversation in terms of a Stochastic Petri Net (SPN). It is well-known that these models are appropriate for almost any sojourn time distribution, usually approximated by a suitable phase-type distribution. Note that the model consists of six states where state "A" (speaker 1 active) coincides with state "$\overline{B}$" (speaker 2 silent) (similarly "$\overline{A}$" coincides with "B"). Remember that the solid transitions allow for immediate firing, whereas the framed ones refer to firing delayed according to timers T_1 to T_4. For a more detailed introduction to SPNs we refer to [1].

Note that this model also complies better with the characterization of a "norm conversation" according to [4], as we can directly adopt the timer values as follows (in msec):

$$T_1 = 1004, \ T_2 = 1587, \ T_3 = 508, \ T_4 = 228 \quad (17)$$

Starting from this parametrization of the norm conversation, we again use τ as a uniform parameter adapting the SPN timers according to

$$T_I = \exp\left(\frac{1}{\kappa\tau} - \frac{1}{\kappa\tau_0}\right) \cdot \overline{T}_I \quad (18)$$

with $\overline{T}_I$ being the timer parametrization for the "norm conversation". Note that Figure 4 refers still to the memoryless case of exponential sojourn time distributions. The corresponding SPNs for general phase-distributions are slightly more complex and subject to current investigations.

6 Summary and Conclusions

This paper proposes a new metric for describing conversational interactivity, using an approach related to thermodynamics in order to derive a scalar equivalent of physical temperature. First results investigating several flavors of our proposal have demonstrated both efficiency and the expressiveness of the method. Therefore, after a much broader validation, this temperature metric will serve as a good candidate for a purely technical description of conversational interactivity.

Acknowledgements

References

[1] F. Bause, P. Kritzinger: *Stochastic Petri Nets.* Vieweg 2002.

[2] P. T. Brady: *A Technique for Investigating On-Off Patterns of Speech.* Bell System Technical Journal, 44(1), pp. 1-22, January 1965.

[3] P. A. Gustavsen, E. Tilley: *Public relations communication through corporate websites: Towards an understanding of the role of interactivity.* PR-ism 1(1), 2003.

[4] ITU-T: *Artificial Conversational Speech.* Rec. P.59, March 1993.

[5] ITU-T: *The E-model, a computational model for use in transmission planning.* Rec. G.107, March 2003.

[6] S. Möller: *Assessment and Prediction of Speech Quality in Telecommunications.* Boston (Kluwer) 2000.

[7] S. M. Ross: *Stochastic Processes.* Wiley 1996.

[8] K. Stowe: *Introduction to Statistical Mechanics and Thermodynamics.* Wiley 1983.

[9] K. Sznajd-Weron, J. Sznajd: *Opinion Evolution in Closed Community.* Int. J. Mod. Phys. C 11: 1157-1166 (2000).