Measuring Non-Spontaneous Interactivity: An Opera-related Case Study

Peter Reichl[1], Marena Balinova[2], Florian Hammer[1]

[1]Forschungszentrum Telekommunikation Wien, Donau-City-Str. 1, A-1220 Vienna, Austria
[2]doremi art management, Seligerstrasse 3, D-91781 Dettenheim, Germany

reichl@ftw.at; marena.balinova@gmx.net; hammer@ftw.at

Abstract

Defining an instrumental concept of interactivity is a fundamental prerequisite for the investigation of interactive multimedia. Unfortunately, this has turned out to be a non-trivial task. Thus, we have recently proposed several approaches for characterizing the interactivity of spontaneous speech conversations. In this paper, we apply our methodology to the non-spontaneous case as represented e.g. by opera duets. Our results demonstrate the validity of the general approach and provide interesting insight into the structure and common patterns of a wide-spread musical form.

1 Introduction

The integration of data and real-time multimedia applications over packet-based transmission technologies like the Internet is one of the central challenges for the future success of telecommunications as such. In this context, both the research community and standardization bodies like ETSI or ITU-T have recently shown increasing interest in characterizing the impact of conversational interactivity on basic network Quality-of-Service (QoS) parameters like packet loss and transmission delay. Unfortunately, it has turned out that already defining the concept of interactivity itself is a non-trivial task [1]. This may explain the lack of a universally accepted metric for interactivity and most recently has motivated a series of papers exploring different approaches to solve this problem. Whereas [2] analyzes several parameters characteristic for conversations and uses them to describe their interactivity, [3] presents an axiomatic approach leading to a metric which closely resembles the thermodynamic concept of temperature. Finally, the entropy-related approach of [4] is especially useful for the case of multi-party conversations.

Conversations play also a vital part in theatre pieces as well as in opera, operetta or musical (up to an extent that e.g. Richard Strauss has termed his last masterpiece "Capriccio" no longer an opera but a "conversation piece for music" [5]). However, there is a severe lack of analysis concerning structure and form of "operatic conversations" (cf. [6]), and to the best of our knowledge it has never been investigated formally to which extent operatic duets or larger ensembles could serve as examples for "non-spontaneous conversations" (which are still representative for structure and pattern of everyday conversations but of course much better reproducible than those).

Therefore, as an initial step into this direction, this paper applies the basic model for conversational interactivity as described in Section 2 to the case of opera duets as a representative example of non-spontaneous conversations. Section 3 presents quantitative results for a selected set of examples, before Section 4 discusses current and future work.

2 Interactivity Metrics

2.1 Conversation/Duet Model

Both everyday conversations and opera duets follow an identical fundamental structure which we can model in accordance with ITU-T Recommendation P.59 [7] by distinguishing four states (see Figure 1). States A and B refer to the situations in which either person/figure A or person/figure B is active alone. State M (mutual silence) represents the case of both persons/figures being silent and state D (double talk) refers to both persons/figures being active simultaneously.

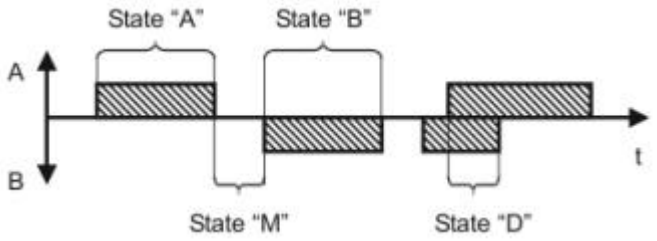

Figure 1. State model of a two-way conversation/duet

2.2 Parametric Conversation/Duet Analysis

As a first approach for analyzing conversational interactivity, [3] has proposed the framework of Parametric Conversation Analysis. It consists of the following basic features and events:

- *Speaker Alternation*: switching event from one speaker/singer to the other, potentially separated by a period of mutual silence.
- *Interruption*: one speaker interrupted by the other who after a phase of double-talk eventually gains the floor.
- *Pause*: mutual silence phase between two talk-spurts of the same speaker.
- *Incision* (*"non-interruptive double talk"*): (short) double-talk phase without speaker alternation.

In contrast to ordinary conversations, "ping-pong" transitions between states *M* and *D* are quite common in a musical context, therefore for our purposes we require one more definition:

- *Parallelism:* joint switching of both participants between mutual-silence and double-talk phases.

Note that extensive periods of parallelism may seriously threaten the comparability of spontaneous and non-spontaneous conversations. This fact has later to be taken into account while choosing candidates for applying the methodology to opera duets.

Based on the features defined above, the following two parameters can be used to describe the interactivity of a given conversation/duet:

- *Speaker Alternation Rate (SAR):* total number of speaker alternations plus interruptions per minute
- *Interruption Rate (IR):* total number of interruptions per minute.

2.3 Temperature Metric

A second approach to formally describe the interactivity of conversations starts from the fact that even if it is hard to describe what interactivity is, there is an intuitive knowledge what it is not. In this sense, summarizing [8], the lack of interactivity is strongly correlated to infinite sojourn times in one of the model states A, B, M or D. In [8] and [3] we have therefore derived formally a metric for conversational interactivity which is based on measuring the mean state sojourn times and comparing them to the respective mean values of an appropriate "reference conversation" (i.e. averages calculated from a large set of representative conversation recordings). Surprisingly enough, it turns out that this approach has a direct analogy in the field of statistical thermodynamics, i.e. the Boltzmann's statistic and the related concept of physical temperature. This fact may be used to calculate a unique scalar metric for interactivity, i.e. the so-called "conversational temperature". Note that the mentioned reference conversation is assigned "room temperature", i.e. 20°; the actual temperature calculation of the conversation/duet is then based on a mean-square approximation algorithm described in [4].

2.4 Results for Speech Conversations

In a first step, we have validated our approach through a series of spontaneous (speech) conversation experiments performed at the Institute of Communications Acoustics at Ruhr-University Bochum, Germany. Here, test persons had to perform different tasks, i.e. either so-called "short conversational tests" like the ordering of pizzas or the booking of flights, or so-called "interactive short conversational tests", including the fast exchange of meteorological data or telephone numbers. These two types of tasks have been deliberately designed in order to provoke a significant difference in the interactivity of the resulting conversations. Table 1 summarizes the results. From this table, we conclude that indeed the task of exchanging weather data has lead uniformly to higher values of the interactivity features than the mimicked pizza service task. Moreover, both tasks have turned out to be less interactive than the reference conversations. One potential explanation refers to the fact that the parameters of P.59 are based on mix of conversations in English, Italian and Japanese, whereas our tests have been performed in German only.

Task/Feature	SAR	IR	Temperature
Pizza Service	19.66	4.28	13.4°
Weather Data	26.04	5.91	14.4°
Reference: P.59	n/a	n/a	20.0°

Table 1: Interactivity Features for Different Conversational Tasks

3 The Interactivity of Opera Duets

In order to investigate further the validity of our approach, we have applied the above measurement methodology for conventional spontaneous speech conversations to a representative set of opera duets that have been chosen according to the following criteria:

- representative for different composition forms, different languages and different epochs/styles
- representative for different emotional content and communication style
- state probability distribution close to voice conversations (especially avoiding the parallelism problem mentioned in section 2.1)

Despite of the huge operatic repertoire, finding good examples complying with these conditions is not easy, especially with respect to the last criterion. For this paper, we have selected the following five duets:

- Rossini, Barber of Seville, Act I, Recitativo Almaviva – Figaro: "*È desso, o pur m'inganno*"
- Beethoven, Fidelio, Act I, Duet Marzelline – Jaquino: "*Jetzt Schätzchen, jetzt sind wir allein*"
- Mascagni, Cavalleria Rusticana, Duet Santuzza – Turiddu: „*Tu qui, Santuzza? – Qui t'aspettavo*"
- Bizet, Carmen, Act IV, Final Duet Carmen – Don José: "*C'est toi? – C'est moi!*"
- Puccini, Tosca, Act II, Final Scene Tosca – Scarpia: "*Tosca, finalmente mia! – Maledetta!!*"

For each example, we have evaluated a reference recording with respect to sojourn times in the four states introduced in section 2 and have calculated their respective median values before applying the algorithms described in section 2. Table 2 presents the resulting values for the interactivity features.

Task/Feature	SAR	IR	Temperature
Barbiere	13.12	2.87	12°
Fidelio	9.66	4.59	14°
Cavalleria	7.49	2.19	11°
Carmen	4.20	1.08	9°
Tosca	21.15	9.40	16°

Table 2: Interactivity Features for Selected Opera Duets

From our results, we conclude that also in the case of opera duets, all three features provide a rather consistent description of the respective interactivity. Additionally, initial user experiments based on listening tests further support the validity of this instrumental approach.

4 Summary and Conclusions

This paper has investigated the interactivity of non-spontaneous conversations. As an especially interesting case, we have chosen to apply interactivity metrics developed in a telecommunications context to opera duets. The encouraging results demonstrate the multi-disciplinary nature of the instrumental framework for describing conversational interactivity. Current and future work focuses on a detailed user study, moreover we will use the insight gained from both the spontaneous and the non-spontaneous case to develop an integrated interactivity metric combining all three features into one appropriate scalar parameter, thus providing a helpful contribution for investigating the impact of interactivity on multimedia transmission.

Acknowledgements

This work has been performed in the framework of the Austrian Kplus Competence Center programme. The authors want to thank G. Kubin and A. Raake for helpful discussions and further support.

References

[1] F. Hammer, P. Reichl, “How to Measure Interactivity in Telecommunications”, Accepted for: 44th FITCE Congress 2005, Vienna, Austria, Sept. 2005.
[2] F. Hammer, P. Reichl, A. Raake, “Elements of Interactivity in Telephone Conversations”, ICSLP’04, Jeju Island, Korea, Oct. 2004.
[3] F. Hammer, P. Reichl, A. Raake, “The Well-Tempered Conversation: Interactivity, Delay and Perceptual VoIP Quality”, Proc. IEEE ICC’05, Seoul, Korea, May 2005.
[4] P. Reichl, G. Kubin, F. Hammer, “A General Temperature Metric Framework for Conversational Interactivity”, Submitted to: ICSLP 2005.
[5] R. Strauss, C. Krauss, “Capriccio. Ein Konversationsstück für Musik in einem Aufzug“, Schott’s Söhne, Mainz, 1942.
[6] G. Altmann: Musikalische Formenlehre. Saur Verlag Munich, 1981.
[7] International Telecommunications Union, “Artificial Conversational Speech”, ITU-T Rec. P.59, March 1993.
[8] P. Reichl, F. Hammer, “Hot Discussions and Frosty Dialogues: Towards a Temperature Metric for Conversational Interactivity”, ICSLP’04, Jeju Island, Korea, Oct. 2004.