

УДК 517.9

Interaction Quality in Human-Human Conversations: Problems and Possible Solutions

Anastasiia V. Spirina*

Eugene S. Semenin†

Institute of Computer Science and Telecommunications
Siberian State Aerospace University
Krasnoyarskiy Rabochiy, 31, Krasnoyarsk, 660014
Russia

Alexander Schmitt‡

Wolfgang Minker§

Institute of Communications Engineering
Ulm University
Albert Einstein-Allee, 43, Ulm, 89081
Germany

Received 10.01.2015, received in revised form 25.02.2015, accepted 23.03.2015

Speech analysis nowadays is widespread. One of its applications is designing Spoken Dialogue Systems, which allow users to interact with computer systems using natural spoken language. The Interaction Quality is a quality metric, which is used in this field to evaluate the quality of interaction between computer and human. It is based on various speech features. The aim of the Interaction Quality model design is to improve Spoken Dialogue Systems by introducing information about Interaction Quality into Spoken Dialogue Modeling. There exists some state-of-the-art related to the Interaction Quality in spoken human-computer communication. In turn, measuring the Interaction Quality for human-human conversation reveals to be an increasingly difficult task. Different types of dialogue exist and for each type the Interaction Quality measure has a different meaning. Furthermore, a specific data corpus is required for modeling the Interaction Quality for each type of dialogue. We describe the idea of developing software tool for semi-automatic dialogue corpus generation, which can help to keep the time for preparing corpora. The Interaction Quality models for human-human conversations can be used for improving Spoken Dialogue Systems in terms of flexibility, human-likeness and user-friendliness. What is more, the results of the Interaction Quality modeling can be useful in the field of manned space exploration for developing systems for automatic monitoring the conditions of the crew of a spaceship, especially for long interplanetary flights. Further development of the work on the Interaction Quality modeling will help to track automatically relationship between crew members on the basis of their speech.

Keywords: interaction quality, human-human conversation, speech analysis, speech/spoken corpus.

Introduction

Speech is the main modality for human communications and the most natural user interface. Using various approaches it is possible to extract different information from speech automatically: textual information, audio/prosodic features of speech, paralinguistic information (such

*s_nastia@mail.ru

†eugenesemenkin@yandex.ru

‡alexander.schmitt@uni-ulm.de

§wolfgang.minker@uni-ulm.de

© Siberian Federal University. All rights reserved

as: gender, emotions, age). Especially speech analysis is widely used in different call-centers. In this case extracted information can help to optimize inbound and outbound calls of a company, to find problems in the customer support (through call-center), to analyze the satisfaction of the callers.

In this paper we have focused on main problems for modeling the Interaction Quality (IQ) for human-human (HH) conversations and suggested some possible solutions. The solution of the existing problems will allow to design the IQ model. Then the further results of the IQ modeling will be useful in the field of manned space exploration for developing systems for automatic monitoring the conditions of the crew of a spaceship, especially for long interplanetary flights. It will help to track automatically relationship between crew members on the basis of their speech.

This paper is organized as follows. The briefly description of the IQ in human-computer (HC) communications is presented in Section 1. Section 2 introduces the IQ in HH conversations, some problems of modeling and possible solutions. Finally, Section 3 presents our conclusion.

1. The interaction quality in spoken human-computer communications

Based on rapid technology development in many fields humans interact with computer systems (spoken dialogue systems (SDS)). There are many computer systems, which help humans: timetable (such as Let's GO: A Spoken Dialog System For The General Public [1]), call routing (Interaction director [2]), different devices voice control and others.

The main task for this system is to help user to achieve his/her aims. In this case the system should understand the user correctly and change its behavior depending on the reactions of the user for a better interaction. That is why it is important to value the interaction quality through the dialogue between a human and a computer system.

Schmitt et al. [3–6] investigated modeling the IQ in HC dialogues. This model is based mostly on user satisfaction. Ideally a model should predict user satisfaction in each point of interaction process and depending on it SDS should change its behavior during the interaction. There exists another work, related to measuring quality of the interaction between humans and an SDS. PARADISE [7] provides quality values on the dialogue level which allows for general optimization of the dialogue in an offline fashion. Unfortunately, this paradigm is not usable for online dialogue optimization where the dialogue system adapts to the current quality of the dialogue.

2. The interaction quality in human-human conversations

The computation of the IQ in HH conversations is important, not only for analyzing the calls in call-centers to solve some problems, but besides to improve SDS, make computer's behavior more humanlike. It means that SDS will be friendlier to users, than now.

Compared with the IQ in spoken HC communication, the computation of the IQ for HH conversation is more complicated.

First of all, we should estimate the speech of both speakers in a dialogue instead of one speaker in HC spoken interaction. It gives us more parameters to compute.

What is more, there are different types of conversations. One of the classifications of dialogue types is presented in [8]. According to this classification dialogue can be: persuasion, inquiry, discovery, negotiation, information-seeking, deliberation, eristic. In addition to this classification the HH conversations can be subdivided in other three big groups:

- task-oriented conversations(the call into the call-center, the discussion to find solution for the problem);

- debates (political debates);
- ordinary conversations (friendly talk and others).

Undoubtedly, there exist software decisions dedicated to task-oriented conversations.

One of such systems belongs to Speech Technology Center (STC). This company specializes on different speech technologies: speech recognition, speech synthesis, speech analysis and other.

One of their solutions is QM analyzer: automatic analysis and evaluation of telephone conversations. This system is an effective toolkit for assessment the work quality of the call-center operators and monitoring customer satisfaction. It includes analysis of different aspects:

- customer's actions;
- voice options;
- speech activity;
- dialogue lexicon;
- dialogue semantic [9].

The window of this system is presented on the Fig. 1.

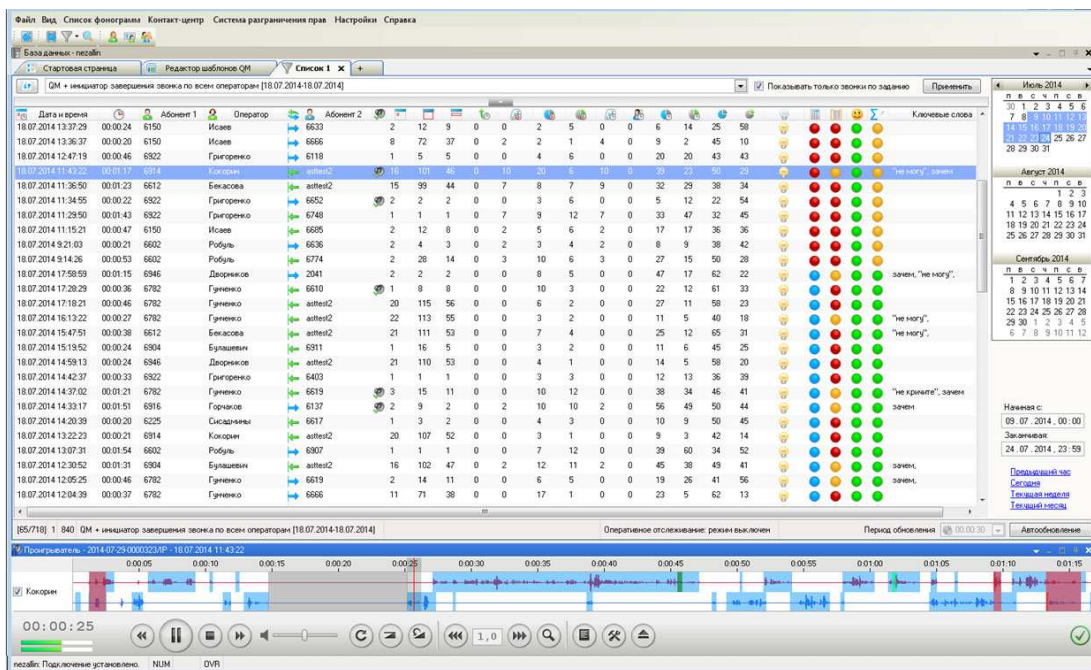


Fig. 1. QM analyzer: automatic analysis and evaluation of telephone conversations

But for designing full model, it is important to research the IQ for other types of conversations.

The first problem for computing the IQ for HH conversation lies in the fact that for each group of conversations the IQ can be interpreted differently. In task-oriented dialogs, concerning calls into the call-centers, the IQ can represent user satisfaction or task completion (achievement the aim of the call). For other types of dialogues, debates and ordinary conversation, the IQ can have more than one interpretation. For example, based on [10] for ordinary conversation the IQ can represent critical discussion of contributions, new ideas from interaction. What is more it can be interpreted as adequacy of conversation, wish of speakers to talk and others. In this case an implementation of a compression of criteria can be useful.

The main problem of the IQ computation in HH conversations is preparing corpus for each type of dialogue.

There are many corpora for speech analysis, but sometimes it is difficult to find suitable corpus for some specific task.

One of the solutions is to develop software for semi-automatic building speech/spoken corpora.

Software for semi-automatic building speech/spoken corpora

The software for semi-automatic building speech/spoken corpora will be useful in fields concerning speech analysis. Researchers need only to collect suitable for research database of audio/video files. Then the program will automatically generate a corpus with selected features. Processing the corpora with this tool will be required the following steps (cf. Fig. 2):

1. Manual selection of the required features from the list for extraction.
2. Automatic extraction of audio information from the video files.
3. Automatic textual information extraction based on the speech recognition and text analysis.
4. Automatic audio/prosodic feature extraction: stress, volume, tempo, pitch, pause, amplitude, quality of voice and others.
5. Automatic paralinguistic information extraction: emotions, gender, age and others.
6. Ability to add to formed corpus expert's labeling.

The formed such a way corpus will contain textual, audio, numerical and expert's data.

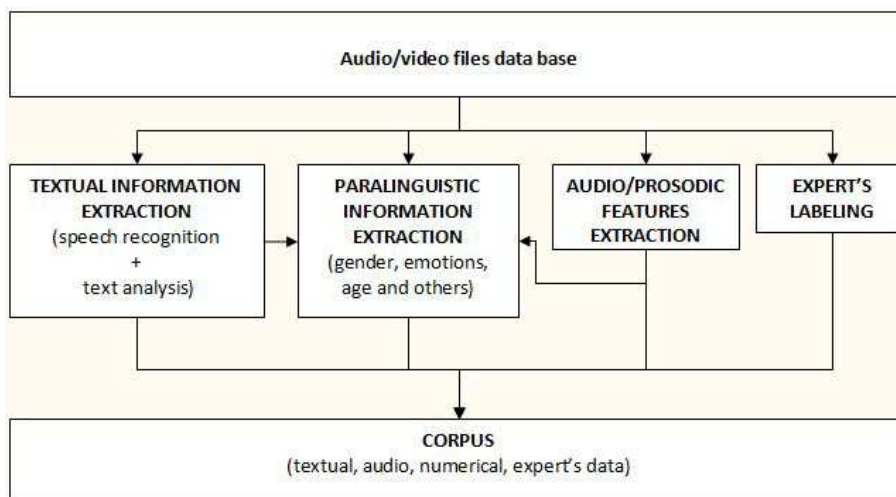


Fig. 2. The scheme for semi-automatic speech/spoken corpus building

Of course, there are different open-source projects, which are able to solve each of the problems: speech recognition, text analysis, emotion recognition, speaker diarization, gender recognition, audio/prosodic features extraction. But to form the corpus you should manually run each of the programs, which requires a lot of time. Using this software for semi-automatic building speech/spoken corpora will help to efficiently generate corpora without major manual intervention.

The proposed program will have modular structure. We plan to integrate in our program open source projects such as LIUM, Praat, openSMILE and others.

LIUM_SPKDIARIZATION is the open source toolkit for diarization: speaker segmentation and clustering [11].

Praat is an open source toolkit for the analysis of speech in phonetics [12].

OpenSMILE (Speech & Music Interpretation by Large Space Extraction) is an open source features extraction utility for automatic speech, music, paralinguistic recognition research [13].

Existing speech/spoken corpora

For each specific problem in the field of speech analysis specific speech/spoken corpus is required. There exist different databases of speech/spoken corpora such as ERLA (European Language Resources Association) [14] and LDC (Linguistic Data Consortium) [15]. It offers corpora for different purposes, such as:

- emotion recognition;
- automatic content extraction;
- discourse analysis;
- information extraction;
- speech recognition;
- language identification;
- speaker identification;
- speaker segmentation;
- speaker verification;
- topic detection and others.

Different corpora consist of different data source:

- video;
- transcribed speech;
- telephone speech;
- telephone conversations;
- telephone speech;
- question-answers;
- microphone conversation;
- microphone speech;
- broadcast conversations and others.

In addition to these there exist some databases of radio program and TV-shows, which can be used in corpus generation. But often information from the existing corpora is not enough for research work. That is why further extension of the corpora is required. The proposed software for semi-automatic building speech/spoken corpora can be used for extension existing corpora for different purpose. The scheme of applying the software for semi-automatic building speech/spoken corpora to existing corpus is presented on the Fig. 3.

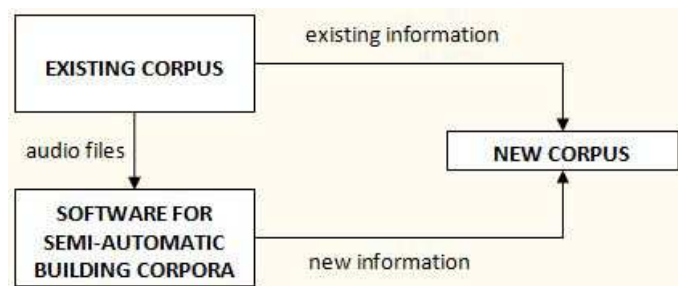


Fig. 3. The scheme of applying the proposed software to existing corpus

3. Conclusion

To sum it up, a software workbench for semi-automatic building corpora will be useful not only for assessment the quality of human-human conversations, it can be used in different lines of investigation in the field of speech analysis, because the main problem in each research in speech analysis is to find appropriate corpus.

Making a corpus for each type of dialogues will solve one of the main problems of the computation of the IQ in HH conversations, which in turn will allow improving of SDS in the future in terms of flexibility, human-likeness and user-friendliness.

What is more, the modular structure of the proposed computer program will allow to expand its functionality depending on the requirements to the certain information in the corpus.

This work was supported by the DAAD (German Academic Exchange Service) together with the Ministry of Education and Science of Russian Federation within Michail Lomonosov Program.

References

- [1] Let's GO: A Spoken Dialog System For The General Public, Available at: <http://www.speech.cs.cmu.edu/letsgo/> (accessed 04.11.2014).
- [2] Interactive intelligence, Available at: <http://www.inin.com/solutions/pages/call-routing-software.aspx>, (accessed 04.11.2014).
- [3] A.Schmitt, B.Schatz, W.Minker, Modeling and predicting quality in spoken human-computer interaction., Proceedings of the SIGDIAL Conference, Portland, Oregon, Association for Computational Linguistics, 2011, 173–184.
- [4] S.Ultes, A.Schmitt, W.Minker, Towards Quality-Adaptive Spoken Dialogue Management, NAACL-HLT Workshop on Future directions and needs in the Spoken Dialog Community: Tools and Data (SDCTD 2012), Association for Computational Linguistics, Montreal, Canada , 2012, 49–52.
- [5] S.Ultes, A.Schmitt, W.Minker, On Quality Ratings for Spoken Dialogue Systems–Experts vs. Users, Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Atlanta, USA, 2013, 569–578.
- [6] S.Ultes, A.Schmitt, W.Minker, A Parameterized and Annotated Corpus of the CMU Let's Go Bus Information System, International Conference on Language Resources and Evaluation, Istanbul, Turkey, 2012, 3369–3373.
- [7] M.Walker, D.Litman, C.A.Kamm, A.Abella, A. PARADISE: a framework for evaluating spoken dialogue agents, Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, 1997, 271–280.
- [8] D.Walton, Types of Dialogue and Burdens of Proof, Computational Models of Argument: Proceedings of COMMA 2010, ed. P. Baroni etc., Amsterdam, IOS Press, 2010, 13–24.
- [9] QM Analyzer, Available at: <http://www.speechpro.ru/product/recognition/asr/qma> (accessed 04.11.2014).
- [10] D.Nandi, S.Chang, S.Balbo, A conceptual framework for assessing interaction quality in online discussion forums, Same places, different spaces, Proceedings of the 26th ASCILITE conference, Auckland, NZ, 2009 , 665–673.

- [11] S.Meignier, T.Merlin, LIUM SpkDiarization: An Open Source Toolkit For Diarization, Proceedings of CMU SPUD Workshop, Dallas, USA, 2010.
- [12] Praat: doing phonetics by computer. Available at: <http://www.fon.hum.uva.nl/praat/>, (accessed 04.11.2014).
- [13] F.Eyben, F.Weninger, F.Gross, B.Schuller, Recent Developments in open SMILE, the Munich Open-Source Multimedia Feature Extractor Proceedings of ACM Multimedia (MM), NY, USA, 2013, 835–838.
- [14] European Language Resources Association, Available at: <http://elra.info/Language-Resources-LRs.html> (accessed 04.11.2014).
- [15] Linguistic Data Consortium, Available at: <https://catalog.ldc.upenn.edu/> (accessed 04.11.2014).

Качество взаимодействия в разговорах типа "человек-человек": проблемы и возможные решения

Анастасия В. Спирина
Александр Шмитт
Евгений С. Семенкин
Вольфганг Минкер

Речевой анализ — одно из быстро развивающихся направлений, обусловленное развитием техники и технологий и, как следствие, появлением необходимости анализа большого объема речевой информации. Одной из областей, в которых применяется речевой анализ, является проектирование речевых диалоговых систем, которые позволяют пользователю взаимодействовать с компьютерными системами на естественном языке. Interaction Quality — это метрика качества, которая используется для оценки качества взаимодействия между компьютером и человеком. Она базируется на различных характеристиках речи. Главной целью разработки модели качества взаимодействия стало улучшение речевых диалоговых систем путем внесения информации о качестве взаимодействия в процесс моделирования диалога. Существует ряд работ, посвященных моделированию качества взаимодействия в речевых диалогах между человеком и компьютером. Оценка качества взаимодействия разговора между людьми гораздо более сложная задача. Существуют различные типы диалогов. Для каждого из них под качеством взаимодействия будут подразумеваться разные понятия. Более того, при моделировании качества взаимодействия для каждого типа диалога требуется свой речевой корпус данных. Подготовка корпуса занимает достаточно много времени. В данной статье мы описываем идею программного инструментария для полуавтоматического создания корпусов, который позволит сэкономить время при создании корпуса. Модель качества взаимодействия для разговоров между людьми поможет в дальнейшем сделать поведение компьютера в речевых диалоговых системах более гибким, похожим на человека и более дружественным. Кроме того, результаты моделирования качества взаимодействия могут использоваться в области пилотируемой космонавтики для разработки систем автоматического мониторинга состояния экипажа космического корабля. Такие системы особенно важны для долгих межпланетных перелетов. Дальнейшее развитие данной разработки позволит автоматически отслеживать отношения между членами экипажа на основе их речи.

Ключевые слова: качество взаимодействия, диалог типа "человек-человек", речевая аналитика, речевой корпус.