V. DISCUSSION

In the previous section, we briefly noted that the appropriateness of the proposed architecture is reflected through the level of its generalizability. It should be noted that the adoption of the proposed architecture does not change the fact that modules of a conversational agent are more often than not developed outside an architectural context. For example, a more detailed discussion on architecture-agnosticity of the state-of-the-art speech recognition modules is provided in [11]. Thus, in this paper, we have intentionally avoided discussing in more detail the design of the modules of the prototype conversational agent (for these details, the reader may consult the given references). However, it is precisely the feature of architecture-agnosticity that allows us to incorporate these modules in an alternative, more general architecture.

The appropriateness of the proposed architecture can also be considered with respect to its scalability and the possibility to support multimodal human-machine interaction. Generally speaking, the scalability of the architecture and a potential multimodality of a conversational agent are simply based on the facts that the design of the architecture is independent of the design of assigned modules, and that the number of modules is arbitrary.

Finally, practical application of this architecture indicates that it should be optimized for conversational agents that incorporate modules requiring intense real-time communication. For example, a module for image processing or a module that controls movements of a robot may generate data on a millisecond-level, in contrast to a speech recognition module that generates data significantly less intensively, as the spoken natural language human-machine dialogue evolves. On the other hand, the tasks of a module for dialogue management include, inter alia, context-dependent interpretation of the data and adaptation of dialogue strategy. If this module tries to processes all data generated in the scope of intense real-time communication in an unselective manner, it may cause congestion.

This problem is not exclusively related to the proposed architecture, but is rather general. Still, in the context of this work, it can be addressed at two levels. At the level of the architecture, this problem can be addressed in such a way that each module can set an acceptance priority for each interaction event type. At the level of a module, these priorities can be dynamically adapted, i.e., according to the current load and interaction context. In this way, a module can dynamically reduce or completely stop the inflow of data. These considerations will be part of future work.

VI. CONCLUSION

This paper introduced a novel modal architecture of a conversational robotic agent and illustrated it for the assistive humanoid robot MARKO. Two important characteristics of this architecture are that (i) it does not depend on specification requirements for a conversational agent, and (ii) its design is separated from the design of modules contained in a conversational agent. Thus, the proposed architecture is general to the extent that it can be applied for a wide range of conversational agents, and modular to the extent that it allows for registration of modules in a general manner, irrespective of their specification requirements and implementation details.

ACKNOWLEDGMENT

The presented study was sponsored by the Ministry of Education, Science and Technological Development of the Republic of Serbia (research grants III44008 and TR32035), and by the intergovernmental network EUREKA (research grant E!9944). The responsibility for the content of this article lies with the authors.

REFERENCES

[1] B. Borovac, M. Gnjatović, S. Savić, M. Raković, M. Nikolić, “Human-like Robot MARKO in the Rehabilitation of Children with Cerebral Palsy”, in: *New Trends in Medical and Service Robots. Assistive, Surgical and Educational Robotics*, Vol. 38 of the series Mechanisms and Machine Science, Springer International Publishing, pp. 191-203, 2016.

[2] E. Gamma, R. Helm, R. Johnson, J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison-Wesley, 1995.

[3] M. Gnjatović, “Therapist-Centered Design of a Robot's Dialogue Behavior”, *Cognitive Computation*, Vol. 6, No. 4, pp. 775-788, 2014.

[4] M. Gnjatović, B. Borovac, “Toward Conscious-Like Conversational Agents”, In: *Toward Robotic Socially Believable Behaving Systems, Volume II - Modeling Social Signals*, A. Esposito, L.C. Jain (eds.), volume 106 of the series Intelligent Systems Reference Library, Springer, pp. 23-45, 2016.

[5] M. Gnjatović, V. Delić, “Cognitively-inspired representational approach to meaning in machine dialogue”, *Knowledge-Based Systems*, Vol. 71, pp. 25-33, 2014.

[6] M. Gnjatović, J. Tasevski, M. Nikolić, D. Mišković, B. Borovac, V. Delić, „Adaptive Multimodal Interaction with Industrial Robot“, in: Proceedings of *2012 IEEE 10th Jubilee International Symposium on Inteligent System and Informatics (SISY)*, Subotica, Serbia, pp. 2012.

[7] N. Hawes, J.L. Wyatt, M. Sridharan, H. Jacobsson, R. Dearden, A. Sloman, G.-J.Kruijff (2010) “Architecture and Representations”, in H.I. Christensen, G.-J. Kruijff, J.L. Wyatt (eds), *Cognitive Systems*, Springer Berlin Heidelberg, pp 51-93, 2010.

[8] N. Jakovljević, D. Misković, M. Janev, D. Pekar, A Decoder for Large Vocabulary Speech Recognition. Proc. of Int. Conf. on Systems, Signals and Image Processing (IWSSIP), Sarajevo; 2011. p. 1--4.

[9] K. Jokinen, M. McTear, *Spoken Dialogue Systems*, Synthesis Lectures on Human Language Technologies. Morgan and Claypool. 2009.

[10] D. Jurafsky, J.H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, second edition, Prentice-Hall. 2009.

[11] D. Mišković, M. Gnjatović, P. Štrbac, B. Trenkić, N. Jakovljević, V. Delić, “Hybrid Methodological Approach to Context-Dependent Speech Recognition”, in *International Journal of Advanced Robotic Systems*, Vol. 14, No. 1, 12 pages, no pagination, 2017.

[12] M. Sečujski, V. Delić, D. Pekar, R. Obradović, D. Knežević, “An Overview of the AlfaNum Text-to-Speech Synthesis System”, in Proceeding of the *XII International Conference ”Speech and Computer” (SPECOM’2007)*, 5 pages, no pagination, Moscow State Linguistic University, Moscow, Russia, 2007.