1. INTRODUCTION

The modality of speech is widely considered to be a promising way to interact with systems in pervasive environments (Sawhney & Schmandt, 2000; Turunen, 2004; Ranganathan, Chetan, Al-Muhtadi, Campbell, & Mickunas, 2005). Being surrounded by a wealth of information processing entities leaves a user with the problem on how to interact with them. While devices can employ dedicated interaction panels (i.e., touch interfaces and graphical displays), such interfaces still require the user to step up to them or carry a mobile device. With the modality of speech comes the promise of casual, effortless and hands-free interaction, suitable to command all the functionality of a pervasive environment.

Apart from the actual speech recognition technology, such an interaction paradigm needs a way to formally describe the spoken dialogs between the system and the user: VoiceXML is one such dialog description language. As part of the standardization efforts of the W3C for multimodal applications, version 2.1 of the language was given recommendation status in June 2007 with version 3.0 being underway. But with its roots in telephony applications and interactive voice response systems, common requirements regarding an application in pervasive environments remain unsatisfied.

Improving the integration of dialogs modeled in VoiceXML with a pervasive environment is, foremost, to extend an interpreters capabilities to exchange information with other systems in the environment and to react accordingly. The illustration in Figure 1 classifies existing VoiceXML language features within a general
push/pull scheme for information exchange between a VoiceXML interpreter and external systems.

While there is language support to pull information into a running VoiceXML session and eventually push gathered information to external systems, the defined semantics exhibit behavior well suited for telephony applications but are cumbersome when applied to pervasive environments. Other communication schemes like pushing information into a running VoiceXML session and pulling information from a running VoiceXML session have no language support at all.

We identified four areas with regard to the communication schemes, where VoiceXML needs to be extended to better support dialogs in pervasive environments:

1. Enable other systems to push information into a running VoiceXML session and allow the interpreter to adapt the dialog.
2. Continuously reflect current values in the object models representing external systems in a VoiceXML session, not the fixed values from the time when a pull was performed.
3. Allow the VoiceXML interpreter to push gathered information to other systems without losing the state of the dialog.
4. Enable other systems to pull information from a running VoiceXML session without waiting for the session to push its information.

In this article, we will describe our work with regard to the first two problems in the context of the JVoiceXML interpreter (http://jvoicexml.sourceforge.net). We consider them to be more important than the last two which we will address in upcoming work.

The rest of this article is organized as follows. After the presentation of related work in Section 2, we give a condensed view of the VoiceXML2.1 standard and its expressiveness in Section 3 and outline core concepts to subsequently identify the possibilities to extend the language with our approach in Section 4. The article concludes with an outlook to VoiceXML3.0 in Section 5 and some subsequent conclusions.

In previous publications, we already described our work of using the JSAPI2 (http://jcp.org/en/jsr/detail?id=113) layer beneath our JVoiceXML implementation to realize a distributed voice user interface incorporating audio streaming to and from mobile devices (Schnelle-Walka & Radomski, 2010; Radomski & Schnelle-Walka, 2010).

2. RELATED WORK

The inability of VoiceXML to react to external events was noticed as early as 2001 for the
Applying Metaheuristics to Minimize Work-Related Musculoskeletal Disorders
[www.igi-global.com/article/applying-metaheuristics-to-minimize-work-related-musculoskeletal-disorders/198991?camid=4v1a](www.igi-global.com/article/applying-metaheuristics-to-minimize-work-related-musculoskeletal-disorders/198991?camid=4v1a)

Automatic Cognitive Load Classification Using High-Frequency Interaction Events: An Exploratory Study
[www.igi-global.com/article/automatic-cognitive-load-classification-using-high-frequency-interaction-events/80412?camid=4v1a](www.igi-global.com/article/automatic-cognitive-load-classification-using-high-frequency-interaction-events/80412?camid=4v1a)