Decalog 2007: Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue, pages 183–184.
Trento, Italy, 30 May – 1 June 2007. Edited by Ron Artstein and Laure Vieu.

Enhancing System Communication through Synthetic Characters

Elena Not, Koray Balci, Fabio Pianesi and Massimo Zancanaro
Bruno Kessler Foundation (FBK-Irst)
Via Sommarive,18
38050 Povo-Trento, Italy
{not,balci,pianesi,zancana}@itc.it

Abstract

Synthetic characters are an effective modality to convey messages to the user, provide visual feedback about the system internal understanding of the communication, and engage the user in the dialogue through emotional involvement. We propose SMIL-AGENT as a representation and scripting language for synthetic characters, which abstracts away from the specific implementation and context of use of the character. SMIL-AGENT has been defined starting from SMIL 0.1 standard specification and aims at providing a high-level standardized language for presentations by different synthetic agents within diverse communication and application contexts.

1 Introduction

Synthetic characters are often integrated in multimodal interfaces as an effective modality to convey messages to the user. They provide visual feedback about the system internal understanding of the communication and engage the user in the dialogue through emotional involvement. However, avatars should not be considered as an indivisible modality, but as the synergic contribution of different communication channels that, properly synchronized, generate an overall communication performance: characters can emit voice and sounds, animate speech with lips and facial expressions, move eyes and body parts to realize gestures, express emotions, perform actions, sign a message for a deaf companion, display listening or thinking postures, and so on.

In this paper, we present SMIL-AGENT, a representation and scripting language that is intended to play as a sort of SMIL dialect for the specification of information presentations by a synthetic agent (Not et al., 2005).

2 SMIL-AGENT

With respect to other existing scripting languages (e.g., CML and AML (Arafa et al., 2004), APML (De Carolis et al., 2002), MPML), SMIL-AGENT pushes further the idea of having a separate representation for the various communication modalities of a synthetic character (e.g., voice, speech animation, sign animation, facial expressions, gestures,…) and their explicit interleaving in the presentation performance. Furthermore, SMIL-AGENT explicitly abstracts away from all data related to the dialogue management and the integration of the agent within larger multimodal presentations, thus assuring the portability of the language (and of the synthetic characters supporting it) to different task and application contexts.
The SMIL-AGENT formalism is certainly less compact and less discourse-oriented than, for example APML (see figure 1 for a sample presentation script written in SMIL-AGENT).

```
<body>
   <par system-language="english">
        <speech channel="alice-voice" affect="sorry-for"
                type="inform" id="say-suffering-angina">
            <mark id="*1*"/>I'm sorry to tell you that you have been
            diagnosed as suffering from <mark id="*2*"/> what we call
            angina pectoris, <mark id="*3*"/> which appears to be mild.
        </speech>
        <seq channel="alice-face" >
             <speech-animation affect="sorry-for"
                               content="say-suffering-angina"
                               end="*3*" intensity="1.5"/>
             <speech-animation affect="positive"
                               content="say-suffering-angina"
                               fill="freeze"/>
        </seq>
        <action channel="alice-right-hand" action-type="pointing"
                content="say-suffering-angina" begin="*2" end="*3">
             <param>bust</param>
        </action>
   </par>
   ...
</body>
```

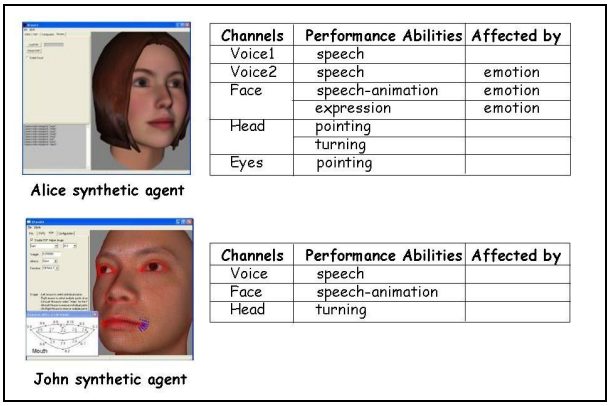
Figure 1. Sample SMIL-AGENT script

As an advantage, it allows plenty of flexibility in expressing which channel should realize a certain performance directive. For example, given a synthetic agent with sophisticated control of body

motion and various channels corresponding to different body parts, a pointing in the direction of the character's bust could be realized by a hand, a finger, a hand plus the head in synchronization, etc…, according to which channel is specified in the <action> element of the script (in the script in Figure 1, for example, the agent's right hand is used). Furthermore, alternative voices could be easily selected at different stages of the presentation, or the face could support a wider set of emotions than voice.

2.1 Extending the language

SMIL-AGENT can easily be used with synthetic characters with different levels of sophistication (figure 2 shows two sample synthetic faces with different communication abilities [1]). The language formal syntax specification defines a separate language partition of attribute values that can be extended by expert authors to list the actual communicative channels and performance abilities supported by a certain synthetic agent. In practice, this is realized by a separate dtd file collecting the list of possible values for: (i) available types of *communicative channels* (e.g., voice, face, eyes, mouth, body, arm,…); (ii) supported *performance abilities* (e.g., verbal and animation abilities, emotions, speech acts, actions, languages); (iii) features that can be tested to include optional parts in the scripts.



Channels	Performance Abilities	Affected by
Voice1	speech	
Voice2	speech	emotion
Face	speech-animation	emotion
	expression	emotion
Head	pointing	
	turning	
Eyes	pointing	

Channels	Performance Abilities	Affected by
Voice	speech	
Face	speech-animation	
Head	turning	

Figure 2. Sample agents with different communication abilities

2.2 Playing scripts

At our institute, a SMIL-AGENT player has been implemented for MPEG-4 based synthetic faces which support: speech, speech-animation, affective facial expressions, gestures, head movements. As shown in Figure 3, for the sake of modularity, the player includes a core processing submodule to which different synthesizers (to get different languages or voice quality) and facial animation players can be plugged in (in the figure, the XfacePlayer and LUCIA[2] players are taken as examples). Visual speech, emotions and expressions are treated as separate channels where the timing is driven by the visual speech to be synchronized with the audio. For each channel, a sequence of morph targets (or FAPs) is created and then blended. An authoring tool for SMIL-AGENT scripts is also under development.

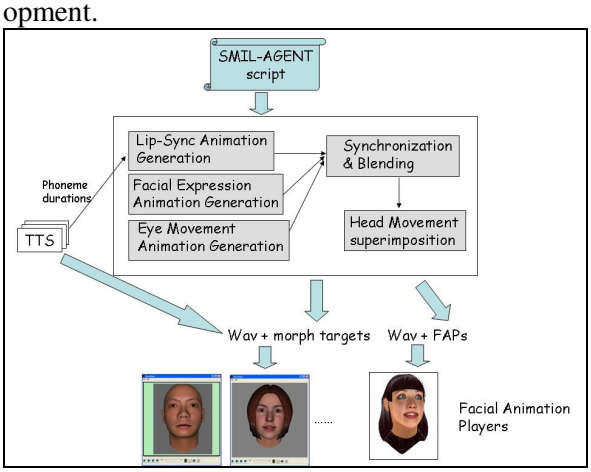


Figure 3. Sample processing of SMIL-AGENT scripts for MPEG-4 based synthetic faces

Reference

Arafa, Y., Kamyab, K. and Mamdani, E. Toward a Unified Scripting Language: Lessons Learned from Developing CML and AML. In H. Prendinger and M. Ishizuka (eds.) *Life-Like Characters. Tools, Affective Functions, and Applications*. Springer-Verlag, 2004, 39-63.

Balci, K. Xface: MPEG-4 based open source toolkit for 3d facial animation. In *Proceedings of AVI04, Working Conference on Advanced Visual Interfaces*, Gallipoli, Italy, 25-28 May, 2004.

De Carolis, B., Carofiglio, V., Bilvi, M. and Pelachaud, C. APML, a Markup Language for Believable Behavior Generation. In *Proceedings of the Workshop on "Embodied conversational agents – let's specify and evaluate them!"* at AAMAS02, 2002

Elena Not, Koray Balci, Fabio Pianesi and Massimo Zancanaro "Synthetic Characters as Multichannel Interfaces" In *Proceedings of ICMI05, Seventh International Conference on Multimodal Interfaces*, Trento, October 3-7, 2005

[1] These faces have been developed with Xface, a set of open tools for the creation of MPEG-4 based 3D Talking Heads (Balci, 2004)).

[2] http://www.pd.istc.cnr.it/LUCIA/home/default.htm.