

Providing Multimodal User Interaction in NCL

Fábio Barreto
MídiaCom Lab - UFF
UNILASALLE-RJ
fbarreto@midia.com.uff.br

Eyre Brasil B. Montevecchi
MídiaCom Lab - UFF
eyrebrasil@midia.com.uff.br

Raphael Abreu
MídiaCom Lab - UFF
raphael.abreu@midia.com.uff.br

Joel A. F. dos Santos
CEFET/RJ
jsantos@eic.cefet-rj.br

Débora C. Muchaluat-Saade
MídiaCom Lab - UFF
debora@midia.com.uff.br

ABSTRACT

This proposal consists of extending NCL events and connectors to provide multimodal user interaction in NCL, creating new types of events such as touch, motion, eyeMotion, pointer, voiceRecognition, gestureRecognition and faceRecognition. New predefined roles, such as onTouch, onMotion, onEyeMotion, onPointer, onVoiceRecognition, onFaceRecognition and onFaceRecognition are provided to express connector conditions.

KEYWORDS

Multimodal User Interaction, NCL, IoT, authoring

1 BACKGROUND

Avanços nas tecnologias de reconhecimento (e.g. reconhecimento de fala ou gesto), deram origem a um tipo de interface denominada MUI (*Multimodal User Interface*). Esse tipo de interface processa duas ou mais modalidades combinadas de entradas de usuário (e.g. fala, toque, gesto, olhar, movimentos da cabeça e do corpo), capturadas por dispositivos de entrada ou sensores, de maneira coordenada com as modalidades de saída, que correspondem a um estímulo para os sentidos humanos (i.e. audição, olfato, tato, paladar ou visão). Tornar as interações do usuário multimodais permite contemplar várias classes e necessidades, uma vez que esses usuários podem utilizar diversos meios de interação com a aplicação.

2 PROPOSAL

A linguagem NCL ainda não possui suporte à multimodalidade, apesar de propostas anteriores sugerirem tal facilidade ([2, 3]). Porém, o modelo NCM, no qual NCL se baseia, já prevê diversos eventos relacionados com a interação do usuário e também atributos que são alterados com a ocorrência destes eventos. Conforme [4], os eventos podem ser dos seguintes tipos: apresentação, composição, seleção, superposição, arraste, foco e atribuição.

Para que seja possível a multimodalidade em NCL, esta contribuição consiste em propor novos tipos de eventos e papéis de conectores NCM, tais como: *touch*, *motion*, *eyeMotion*, *pointer*, *voiceRecognition*, *gestureRecognition* ou *faceRecognition*. Eles representam respectivamente a interação por toque, movimento do corpo, movimento dos olhos, apontamento, reconhecimento de voz, reconhecimento de gesto e reconhecimento de expressão facial. Tais eventos

In: Future of Interactive Television Workshop (V WTVDI), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2019.
ISSN 2596-1683

possuem o atributo *key*, que especifica o que foi reconhecido, de maneira análoga ao que ocorre com o tipo de evento seleção, onde o atributo *key* indica a tecla que foi pressionada pelo usuário. Os demais eventos utilizam o atributo *key* para identificar detalhes sobre a interação multimodal. Por exemplo, para um evento de *gestureRecognition*, o valor do atributo *key* igual a *handUp* indica a ocorrência de um gesto de levantar a mão. A definição da semântica dos valores do atributo *key*, ou seja, dos tipos de gestos, comandos de voz, expressões faciais, etc., é dependente do tipo de evento em questão. Na linguagem NCL, a interação multimodal é expressa com adição de eventos e novos papéis para condições simples de conectores NCL (atributo *role* do elemento *<simpleCondition>*). A Tabela 1 apresenta os papéis relacionados a eventos de interatividade multimodal propostos e a semântica utilizada para o valor do atributo *key*.

3 USE CASE(S)

O caso de uso que ilustra a utilização da multimodalidade em NCL é uma aplicação com objetivo de ajudar na realização de terapias de Integração Sensorial em indivíduos com TEA (Transtorno do Espectro Autista). Essa consiste em orquestrar um ambiente imersivo onde estaria inserido o indivíduo (João) e a sua terapeuta (Marina). São proporcionadas várias experiências sensoriais controladas pelo terapeuta. O suporte a múltiplos usuários em NCL é proposto em outra contribuição [1]. A Listagem 1 contém a especificação de parte do funcionamento desse ambiente, onde o vídeo (*trafficVideo*) pode ser interrompido quando o paciente demonstrar insatisfação através de seu semblante (papel *onFaceRecognition* com expressão facial do tipo triste (*key="SAD"*)).

```
1 <userBase>
2   <userProfile id="paciente" max="1" description="
3     autista.rdf"/>
4   <userProfile id="terapeuta" max="1"
5     description="TO.rdf"/>
6   <userAgent id="joao" profile="paciente"/>
7   <userAgent id="marina" profile="terapeuta"/>
8 </userBase>
9 <causalConnector id="onFaceRecognitionStop">
10  <connectorParam name="face"/>
11  <connectorParam name="owner"/>
12  <simpleCondition role="onFaceRecognition" key="
13    $face" owner="$owner"/>
14  <simpleAction role="stop"/>
15 </causalConnector>
```

Table 1: Papéis propostos para contemplar as interações multimodais

Nome do Papel	Descrição	Atributo <i>key</i>
<i>onVoiceRecognition</i>	Quando algo dito pelo usuário for reconhecido enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	trecho de voz reconhecido
<i>onFaceRecognition</i>	Quando alguma expressão facial do usuário for reconhecida enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	tipo da expressão facial do usuário
<i>onGestureRecognition</i>	Quando algum gesto feito pelo usuário for reconhecido enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	tipo do gesto do usuário
<i>onEyeMotion</i>	Quando algum movimento dos olhos do usuário for reconhecido enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	direção do movimento dos olhos do usuário
<i>onMotion</i>	Quando algum movimento corporal feito pelo usuário for reconhecido enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	tipo do movimento do corpo do usuário
<i>onTouch</i>	Quando alguma interação por toque feita pelo usuário for reconhecida enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	-
<i>onPointer</i>	Quando um apontamento realizado pelo usuário for reconhecido enquanto o objeto associado a esse papel estiver sendo apresentado ou com foco.	-

```

15 ...
16 <link xconnector="onFaceRecognitionStop">
17   <bind role="onFaceRecognition" component="
18     trafficVideo"/>
19     <bindParam name="face" value="SAD"/>
20     <bindParam name="owner" value="joao"/>
21   <bind role="stop" component="trafficVideo"/>
  </link>

```

Listagem 1: Especificação de exemplo multimodal

ACKNOWLEDGMENTS

Os autores gostariam de agradecer a CAPES, CNPq e FAPERJ pelo apoio financeiro parcial deste trabalho.

REFERENCES

- [1] Fabio Barreto, Eyre Montevocchi, Raphael Abreu, Joel Santos, and Debora Muchalut-saade. 2019. Providing multi-user in NCL with userAgent and User-Profile. In *V Future of Interactive Television Workshop* (2019).
- [2] Álan Lívio Vasconcelos Guedes, Roberto Gerson de Albuquerque Azevedo, and Simone Diniz Junqueira Barbosa. 2017. Extending multimedia languages to support multimodal user interactions. *Multimedia Tools and Applications* 76, 4 (01 Feb 2017), 5691–5720. <https://doi.org/10.1007/s11042-016-3846-8>
- [3] C. A. S. Santos, A. N. R. Neto, and E. B. Saleme. 2015. An Event Driven Approach for Integrating Multi-sensory Effects to Interactive Environments. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*. 981–986. <https://doi.org/10.1109/SMC.2015.178>
- [4] Luiz Fernando Gomes Soares and Rogério Ferreira Rodrigues. 2005. Nested context model 3.0: Part 1–ncm core. *Relatório Técnico de Pesquisa da série de Monografias do Departamento de Informática da PUC-Rio* 12 (2005).