Multimodal output

Objectives

Human-machine communication can be defined as a framework of relationships between three actors: the user, the application and the interface. The exchange of information between the user and the machine constitutes the medium of communication. This leads us to consider the electronic document as an object of communication between an author and one or more readers. On this object, various processes such as comprehension, retrieval or appropriation of the text are brought into play through an interface. A structured electronic document can be splitted into:

- its textual content,
- and its typo-dimensional attributes which represent the material formatting of the text, such as the "title" style.

The objective here is to study one of the properties of multimodality CARE - Complementarity, Assignment, Redundancy and Equivalence (defined by (Coutaz, 1994) for blind users.
In this tutorial, we will study the "equivalence" property in the context of non-visual interaction. We call equivalence the state in which the results or means of expression are identical. Three types of equivalence can be defined:

- **equivalence of effect**, which is the fact of obtaining the same result by using different modes of expression and effectors. On the input side, equivalence of effect refers to the fact of obtaining the same result at the task level - action - whether by using keys or oral input, for example. On the output side, the semantics of the message generated by the system is the same: it is produced on different receivers and interpreted in different ways according to the perceptual abilities of the user. However, the equivalence between two modalities does not mean that the information transmitted by these modalities is precisely the same
- **functional equivalence** between means of expression in terms of accuracy and ease of use for users. This equivalence is based on the user's senses. The criteria that define it are the adequacy of the meaning to the task, the ease of use, the completeness between a visual representation and one or more substitute representations. It can be measured in terms of cognitive cost.
- to which we can add **the equivalence of use** between several users.

The objective is **to develop a multimodal fission engine for multimodal presentation of texts.** You will have to code:

- **concurrent multimodality** where the output media are used in parallel in a disjoint manner. The redundant information will be limited for our application to the presentation of typo-dimensional attributes. Thus, these attributes are translated on several media at the same time, for example, displayed on a Braille display (simulated) and spoken (corresponds to the effect equivalence in CARE)
- **synergistic multimodality**, for which you can:
    - **have the typographical attributes stated orally:** for example ("**level 1 title**: Toulouse, Ville Rose"). This is the simplest solution, but it is the one that requires the most cognitive effort from the user, who must memorize the text by mentally associating its possible attributes.
    - **vary the prosodic parameters** of the speech synthesis (*rate*, *elocution* and *volume*) according to the attribute in question. For example, the attribute **"bold"**, which may denote a sentence with strong semantics, will have to be spoken at a slower speed and a high volume.
    - Or **use a** simulated **Braille display** (for the blind) on which the typographic attributes in play are displayed while the synthesis renders the text.

You can use the document available here (**https://github.com/truillet/upssitech/blob/master/SRI/3A/IHM/TP/Code/Toulouse.html**)for your tests

Available tools:

- **ppilot5,** ivy text-to-speech agent (supports SSML FORMAT **(https://www.w3.org/TR/speech-synthesis11**)
- **Braille_display,** ivy agent simulating a 10-dot braille display
- The ivy middleware

You can code your engine in any language you like.
The choice of the desired multimodality as well as the *play / pause / return to the beginning of the playback* will be implemented by the engine L

Links

- Coutaz J., Nigay L., Les propriétés "CARE" dans les interfaces multimodales. IHM'94, Lille, 8 & 9 décembre 1994, pp. 7-14
- Schnelle-Walka, Dirk; Radomski, Stefan; Mühlhäuser, Max (2014): Multimodal Fusion and Fission within W3C Standards for Nonverbal Communication with Blind Persons, pp. 209-213, Springer, Computers Helping People with Special Needs, 14th International Conference on Computers Helping People with Special Needs, ISBN 978-3-319-08595-1 (**https://www.icvr.ethz.ch/research/projects/closed/dach/publications/fission_icchp_2014.pdf**)