

Multimodalité en sortie

Objectifs

La communication humain-machine peut se définir comme un cadre de relations entre trois acteurs : l'utilisateur, l'application et l'interface. Les échanges d'informations entre l'utilisateur et la machine constituent le support de la communication. Ceci nous amène à considérer le document électronique comme un objet de communication entre un auteur et un ou plusieurs lecteurs. Sur cet objet, différents processus tels que la compréhension, le repérage ou l'appropriation du texte sont mis en jeu au travers d'une interface. Un document électronique structuré se décompose en :

- son contenu textuel,
- et ses attributs typo-dimensionnels qui représentent la mise en forme matérielle du texte comme par exemple le style "titre".

Ces derniers sont porteurs de sens en sus du contenu du document proprement dit.

L'objectif est ici d'étudier une de des propriétés de la multimodalité CARE- Complémentarité, Assignment, Redondance et Équivalence (définies par (Coutaz, 1994) pour des usagers non-voyants.

Nous allons durant ce TP étudier « l'équivalence » dans le cadre de l'interaction non-visuelle. Nous appelons équivalence l'état dans lequel les résultats ou les moyens d'expression sont identiques. Trois types d'équivalence peuvent être définies :

- **L'équivalence d'effet** qui est le fait d'obtenir un même résultat en utilisant des modes l'expression, des effecteurs différents. En entrée, l'équivalence d'effet désigne le fait d'obtenir au niveau de la tâche —action— le même résultat que ce soit par l'usage des touches ou de l'entrée orale, par exemple. En sortie, la sémantique du message générée par le système est la même : elle est produite sur différents récepteurs et interprétée de manière différente suivant les capacités perceptives de l'utilisateur. Cependant, l'équivalence entre deux modalités ne signifie pas que l'information transmise par ces modalités soit précisément la même
- **L'équivalence fonctionnelle** entre moyens d'expression quant à leur précision et en facilité d'usage pour les utilisateurs. Cette équivalence est basée sur les sens de l'utilisateur. Les critères qui la définissent sont l'adéquation du sens à la tâche, l'aisance d'usage, la complétude entre une représentation visuelle et une ou des représentation(s) de substitution. Elle peut se mesurer en termes de coût cognitif.
- auxquelles nous pouvons ajouter **L'équivalence d'usage** entre plusieurs utilisateurs.

L'objectif est de **développer un mini-moteur de fission multimodale permettant de présenter de manière multimodale des textes**. Vous aurez à coder :

- la **multimodalité de type concurrent** où l'on utilise en parallèle les médias de sortie de manière disjointe. L'information redondante va être pour notre application limitée à la présentation des attributs typo-dimensionnels. Ainsi, ces attributs sont traduits sur plusieurs médias à la fois, par exemple, affichés sur une plage braille (simulée) et énoncés oralement (correspond à l'équivalence d'effet dans CARE)
- la **multimodalité synergique** pour laquelle vous pourrez :
 - **faire énoncer oralement les attributs typographiques** : par exemple (« **titre de niveau 1** : Toulouse, Ville Rose »). C'est la solution la plus simple mais c'est celle qui sollicite le plus la

charge cognitive de l'utilisateur, celui-ci devant mémoriser le texte en associant mentalement ses éventuels attributs.

- **faire varier les paramètres prosodiques** de la synthèse vocale (*débit, élocution et volume*) d'après l'attribut mis en jeu. Par exemple, l'attribut « **gras** », pouvant dénoter une phrase avec une sémantique forte, il faudra l'énoncer avec une vitesse plus lente et un fort volume.
- Ou enfin **utiliser une plage braille** simulée (pour les non-voyants) sur laquelle les attributs typographiques en jeu sont affichés pendant que la synthèse restitue le texte.

Vous pouvez utiliser le document disponible ici

(<https://github.com/truillet/upssitech/blob/master/SRI/3A/IHM/TP/Code/Toulouse.html>) pour effectuer vos tests

Outils disponibles :

- **ppilot5**, agent ivy de synthèse vocale (prend en compte le format SSML (<https://www.w3.org/TR/speech-synthesis11>))
- **Braille_display**, agent ivy simulant une plage braille 10 points braille
- Le middleware ivy

Vous pouvez coder votre moteur dans le langage que vous voulez.

Le choix de la multimodalité désirée ainsi que les fonctions *play / pause / retour au début de la lecture* seront implémentées par le moteur

Liens

- 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)