



Projet VERBATIM

## Sous-projet 5

# *Expérimentation utilisateurs d'un service multimodal : "Plan resto"*

Version : 2

Date : 15/12/2006

Auteur(s) : Joëlle Blanquet, Jérôme Gatefin (France Telecom)



# Table des matières

<b>1. Introduction</b>	<b>4</b>
<b>2. Objectifs et contexte de l'étude</b>	<b>5</b>
<b>2.1 Objectifs</b>	<b>5</b>
<b>2.2 Choix du service testé</b>	<b>5</b>
<b>2.3 Cadre théorique</b>	<b>6</b>
2.3.1 Définition de la multimodalité	6
2.3.2 Types de coopérations entre modalités	7
2.3.3 Les avatars virtuels	8
<b>3. Méthode</b>	<b>9</b>
<b>3.1 Déroulement de l'étude</b>	<b>9</b>
<b>3.2 Sujets</b>	<b>9</b>
<b>3.3 Matériel</b>	<b>10</b>
3.3.1 Versions multimodales	10
3.3.2 Version graphique	14
3.3.3 Version vocale	15
3.3.4 Enregistrement vidéo	15
3.3.5 Questionnaires et entretiens	15
<b>3.4 Procédure</b>	<b>16</b>
<b>3.5 Scénarii</b>	<b>17</b>
<b>3.6 Recueil des données</b>	<b>17</b>
3.6.1 En termes d'efficacité	18
3.6.2 En termes d'efficacité	18
3.6.3 En termes d'utilisation des modalités	18
3.6.4 En termes de satisfaction	18
<b>3.7 Hypothèses</b>	<b>19</b>
<b>4. Résultats</b>	<b>20</b>
<b>4.1 Données quantitatives</b>	<b>20</b>
4.1.1 Temps mis pour effectuer chaque scénario	20
4.1.2 Nombre d'interactions effectué pour chaque scénario	21
4.1.3 Fréquence d'utilisation des différentes modalités (pour les interfaces multimodales uniquement)	22
4.1.4 Profils d'utilisation des modalités (pour les interfaces multimodales uniquement)	23
4.1.5 Indice de charge mentale	24
4.1.6 Indice de satisfaction	25
<b>4.2 Données qualitatives</b>	<b>25</b>
4.2.1 Concernant la multimodalité en entrée	25
4.2.2 Concernant la multimodalité en sortie	26

<b>4.3</b>	<b>Discussion des résultats.....</b>	<b>28</b>
4.3.1	Concernant l'efficience.....	28
4.3.2	Concernant l'efficacité.....	28
4.3.3	Concernant les modalités utilisées.....	29
4.3.4	Concernant la redondance.....	30
4.3.5	Concernant la satisfaction.....	30
<b>5.</b>	<b>Recommandations.....</b>	<b>32</b>
<b>5.1</b>	<b>Concernant la multimodalité en entrée.....</b>	<b>32</b>
5.1.1	Le choix entre modalités.....	32
5.1.2	La modalité vocale.....	33
<b>5.2</b>	<b>Concernant la multimodalité en sortie.....</b>	<b>33</b>
5.2.1	Complémentarité modale et redondance.....	33
5.2.2	La synthèse vocale.....	34
5.2.3	Feedback.....	34
5.2.4	L'avatar.....	35
<b>6.</b>	<b>Conclusion, perspectives.....</b>	<b>36</b>
<b>7.</b>	<b>Bibliographie.....</b>	<b>37</b>
<b>8.</b>	<b>Annexes.....</b>	<b>38</b>

## 1. Introduction

Le présent document se situe dans le cadre du projet RNRT VERBATIM, dont l'objet est la VERification Biformelle et Automatisation du Test des Interfaces Multimodales.

Le Sous Projet 5 concerne les tests utilisateurs de services multimodaux in situ.

Le Lot 1 est consacré aux aspects théoriques appliqués à l'interaction multimodale.

Les autres lots devaient permettre d'étudier les résultats obtenus par différentes méthodes de tests : d'une part les tests utilisateurs in situ menés par des ergonomes et d'autre part les tests automatiques de recommandations ergonomiques réalisés par des approches formelles.

Les travaux menés dans le SP1 nous ont amené à constater que le domaine de la multimodalité est encore très jeune du point de vue de l'étude des usages réels de sorte que peu de recommandations ergonomiques pour les interfaces multimodales ont pu être examinées par les approches formelles dans les délais initialement prévus. Il a donc été décidé de monter une expérimentation de service multimodal avec des utilisateurs, dans un contexte réaliste, et de façon approfondie, pour pouvoir alimenter les travaux des différents sous-projets.

En conséquence, le présent document remplace les livrables du lot 3.

## 2. Objectifs et contexte de l'étude

### 2.1 Objectifs

---

L'objectif de cette étude est d'élaborer des recommandations ergonomiques concernant les interfaces multimodales afin d'alimenter le projet VERBATIM, dont le but est d'explorer la possibilité de vérifier automatiquement des recommandations ergonomiques multimodales.

En outre, l'étude menée a pour but d'apporter des éléments de réponse aux questions suivantes :

- Les interfaces multimodales sont-elles plus efficiente (en termes de temps mis pour réaliser les différents scénarios) que les interfaces monomodales pour un même service de recherche d'informations ?
- Les interfaces multimodales sont-elles plus efficace (en termes d'économie : nombre d'interactions, indice de charge mentale) que les interfaces monomodales pour un même service de recherche d'informations ?
- Comment les utilisateurs interagissent-ils avec les interfaces multimodales (en termes de modalités utilisées) ?
- Comment les interfaces multimodales sont-elles perçues du point de vue de la satisfaction des utilisateurs par rapport aux interfaces monomodales?
- La présence d'un avatar virtuel affecte-t-elle l'efficacité, l'efficacé, le comportement multimodal et/ou la satisfaction des utilisateurs ?

### 2.2 Choix du service testé

---

La recherche d'un service multimodal suffisamment robuste pour être testable par des utilisateurs en situation réaliste, nous a conduits à choisir le service "Plan Resto". Les autres prototypes disponibles ne permettaient pas d'effectuer des tests auprès d'utilisateurs dans de bonnes conditions notamment à cause de la présence de nombreux bugs.

Nous avons également écarté les prototypes de service multimodaux dans lesquels sont implémentés de la réalité virtuelle. En effet, il n'est pas toujours possible alors de savoir si les résultats obtenus sont imputables à une technologie (la multimodalité), plutôt qu'une autre (la réalité virtuelle).

Plan Resto multimodal était implémenté sur PC (interface web) et était disponible en 2 versions : avec et sans avatar virtuel. Ce service existait également en version uniquement graphique (sur PC) et uniquement vocal (sur téléphone fixe). Le fait de

bénéficier de ces différentes versions a présenté l'avantage de pouvoir tester en parallèle les versions multimodales et monomodales d'un même service.

## **2.3 Cadre théorique**

---

Les interfaces multimodales, dans le domaine des Interactions Homme-Machine permettent aux utilisateurs de pouvoir utiliser plusieurs modalités de communication comme la parole, l'écrit ou encore le geste. L'un des objectifs de ce type d'interface est de se rapprocher de la communication humaine, en essayant de rendre ces interfaces plus naturelles et spontanées dans leur utilisation. La multimodalité peut être présente en entrée du système, en sortie ou les deux. Après une définition de la multimodalité et des différentes coopérations qui peuvent exister entre modalités. Nous parlerons ensuite des avatars virtuels qui peuvent être implémentés dans ce type d'interface et enfin nous exposerons les appuis théoriques pour l'évaluation des interfaces multimodales.

### **2.3.1 Définition de la multimodalité**

Une interface peut être définie comme multimodale, à partir du moment où elle utilise au moins deux modes d'entrées combinés (comme la parole, les mouvements de la tête et du corps, le stylo, tactile, etc.) coordonnés avec les systèmes de sorties multimédias (Oviatt, 2003).

La conception des interfaces multimodales n'a pas comme unique but d'exploiter au mieux les ressources techniques (traitement automatique de la parole, reconnaissance automatique des mouvements, etc.), mais de développer des techniques qui soient conformes aux attributs de la communication humaine. La multimodalité prend son sens aussi bien dans la variété des canaux d'interactions à destination de l'utilisateur (en sortie) que dans les diverses manières d'utiliser ces canaux (en entrée) (Nigay, 2001).

Sharon Oviatt, qui a mené de nombreuses expérimentations dans le domaine de la multimodalité, décrit les avantages des interfaces multimodales dans une de ses publications (Oviatt, 2003). Ainsi, selon cet auteur, les interfaces multimodales permettent à l'Interaction Homme-Machine d'être plus transparente, plus flexible, plus efficiente et plus expressive que les interfaces monomodales traditionnelles (ordinateur muni d'un clavier et d'une souris, téléphone mobile, etc.). De plus, ce type d'interface permet de faciliter l'apprentissage et l'utilisation, et d'augmenter la satisfaction des utilisateurs.

Dans le cadre du projet VERBATIM, l'interface multimodale que nous avons testée permet de trouver un restaurant dans Paris à l'aide de trois critères de recherche et

dispose, pour cela, des modalités d'interaction présentées dans le tableau 1 ci-dessous :

Tableau 1 : Modalités et modes utilisés dans le cadre de cette étude

Modalités	Entrées vers la machine		Sorties vers l'utilisateur	
	Actions utilisateurs	Support matériel	Présentation par la machine	Support matériel
Visuelle	Ecriture	Clavier	Affichage de texte	Ecran
	Clic sur lien ou bouton	Souris	Affichage d'image, activation de bouton...	
Sonore	Parole	Reconnaissance vocale	Synthèse vocale	Haut-parleurs

En entrée (de l'utilisateur vers la machine) : **équivalence des modalités**

- le clavier : pour écrire une requête
- la souris : pour cliquer sur des liens ou sur des boutons
- la parole : pour énoncer oralement une requête qui est ensuite traitée par la reconnaissance vocale

En sortie (de la machine vers l'utilisateur) : **redondance**

- l'écran : pour afficher du texte, des images, des boutons
- les haut-parleurs : pour diffuser la synthèse vocale

### 2.3.2 Types de coopérations entre modalités

Il existe quatre types de coopération entre modalités (d'après Bellik, 1995) :

- **L'équivalence** : fait référence au fait que l'utilisateur ou la machine peut avoir le choix entre plusieurs modalités pour formuler un énoncé particulier. Par exemple, en entrée, l'utilisateur peut avoir le choix entre prononcer le nom de la commande qu'il désire exécuter ou la désigner sur une barre d'outils en utilisant la souris. En sortie, la machine peut par exemple, avertir l'utilisateur de l'arrivée d'un nouveau mail, soit en émettant un message sonore, soit en affichant un message visuel.

- **La redondance** : consiste à transmettre un même énoncé sur différentes modalités. En sortie, la redondance peut être utilisée pour signaler l'importance ou l'urgence d'un message (dans ce cas, la redondance fournit une information supplémentaire)
- **La complémentarité** : consiste à transmettre différents messages représentant les constituants d'un énoncé sur plusieurs modalités. La compréhension de l'énoncé nécessite ici une fusion entre les différents messages transmis à travers les différentes modalités. Ainsi, en entrée, l'utilisateur peut prononcer la phrase "*efface ce fichier*" tout en cliquant avec la souris sur l'icône du fichier en question. En sortie, une synthèse vocale peut, par exemple, expliquer un schéma de géométrie affiché sur l'écran.
- **La spécialisation** : consiste à utiliser toujours la même modalité pour un type d'énoncé particulier et à ne pas l'utiliser pour un autre type d'énoncés (spécialisation exclusive). Par exemple, en entrée et dans une application graphique de dessin, on peut affecter la parole à l'entrée des noms des commandes et la souris à l'entrée des paramètres des commandes. En sortie, par exemple dans une application d'édition de textes pour handicapés visuels, on peut décider d'utiliser l'affichage braille uniquement pour la présentation du texte, les messages d'erreurs et autres "feed-back" étant communiqués à travers la synthèse de parole.

### 2.3.3 Les avatars virtuels

Il est possible de doter les interfaces multimodales d'avatars virtuels. Cependant, le bénéfice apporté par la présence de ces derniers n'est pas totalement démontré. En effet, certaines études suggèrent que les avatars seraient susceptibles d'améliorer l'efficacité d'une application (Beun et al., 2003 ; Ryokai et al., 2003 in Buisine, 2005). Néanmoins l'expérimentation d'André et al. (1998) n'a pas permis de mettre en évidence des différences, en termes de performance, entre une interface avec avatar et sans avatar.

## 3. Méthode

### 3.1 Déroulement de l'étude

---

La première phase de l'étude a consisté à réaliser un état de l'art synthétique, à partir d'études académiques d'une part, et à partir des études France Télécom R&D menées depuis plusieurs années d'autre part. Cette phase a permis d'établir un certain nombre de définitions, de concepts, de questions et d'hypothèses de travail. Pour répondre à ces questions et étudier ces hypothèses, nous avons opté pour une étude comparative du service "Plan resto" selon quatre versions différentes : vocale, graphique, multimodale sans avatar, et multimodale avec avatar.

Nous avons réalisé dans une seconde phase, une expertise ergonomique multi-experts des différentes versions du service "Plan resto", afin d'améliorer autant que possible l'utilisabilité du service et faciliter l'interaction. Ceci a été entrepris afin de réduire les difficultés de l'utilisateur quant à la compréhension du service et pouvoir mieux se centrer sur la comparaison des différents types d'interaction. L'expertise ergonomique a permis de recenser un certain nombre d'éléments de difficulté. Nous avons alors maqueté une nouvelle version de l'interface graphique et proposé aux développeurs un certain nombre d'évolutions afin d'améliorer la lisibilité de l'interface graphique (par exemple, le repérage des différentes zones d'information, des boutons de commandes et des zones de saisie) et la cohérence entre le graphique et le vocal (terminologie homogène). Les discussions avec les différents concepteurs-développeurs ont abouti à une version améliorée et plus cohérente des interfaces graphique et vocale.

La phase suivante a consisté à établir un plan expérimental permettant d'étudier les questions et les hypothèses retenues précédemment. Une fois le plan expérimental défini, il s'est agi de définir précisément la séance de test (scénarii, recueil des données, accueil des sujets, questionnaires, etc.) de réaliser un pré-test afin de vérifier le bon déroulement de la séance, de recruter les sujets, et de faire passer le test à 24 sujets. La dernière phase a été consacrée à l'analyse des résultats et à l'élaboration de recommandations ergonomiques.

### 3.2 Sujets

---

Les 24 personnes (12 hommes et 12 femmes) ayant participé à cette étude étaient des employés de France Télécom R&D, des stagiaires et des personnes faisant partie du réseau de connaissance. L'âge moyen des participants était de 31 ans et 6 mois (écart-type,  $E.T.=10,59$ ). Ils ont été répartis en 4 groupes :

- 6 personnes ont réalisé les tests avec une interface vocale téléphonique (téléphone fixe)
- 6 personnes ont réalisé les tests avec une interface graphique (web) sur PC
- 6 personnes ont réalisé les tests avec une interface multimodale (web) sur PC
- 6 personnes ont réalisé les tests avec une interface multimodale dotée d'un avatar virtuel (web) sur PC

### 3.3 Matériel

---

Le service Plan Resto permet de trouver un restaurant dans Paris intra-muros grâce à trois critères de recherche : le lieu (arrondissement, station de métro ou monument de Paris), le prix et la spécialité culinaire. Néanmoins, il n'est pas possible d'effectuer une recherche à partir d'un nom de restaurant. Pour les besoins des tests, nous avons utilisé quatre versions du service Plan Resto.

#### 3.3.1 Versions multimodales

L'interface à l'écran est décomposée en cinq zones :

- **En haut** se trouve l'historique du dialogue entre l'utilisateur et Plan Resto. Ainsi l'utilisateur peut lire la trace de l'interaction :
  - ce qu'a dit la synthèse vocale avec possibilité de cliquer sur les liens qui lui sont proposés (les propositions et réponses du système).
  - ce qu'a pris en compte la reconnaissance vocale de ce que l'utilisateur a lui-même dit (les questions de l'utilisateur).
- **Au milieu**, un champ de saisie permet à l'utilisateur de taper sa requête et de la valider grâce au bouton "Ok". Sont également présents dans cette zone les boutons "*Nouvelle recherche*", "*Mode d'emploi*" et "*Réinitialiser Plan Resto*" (en cas de bug).
- **En bas à gauche**, une zone est dédiée uniquement à la liste des résultats (liste des restaurants). Chaque nom de restaurant répondant aux critères de l'utilisateur est ainsi affiché et cliquable.
- **En bas au milieu** sont affichés une zone où figure l'avatar (pour la version avec avatar seulement) et un bouton "*Cliquez ici pour parler*" (présent dans la version avec avatar et sans avatar) qui permet à l'utilisateur d'interagir vocalement avec Plan Resto.
- **En bas à droite**, une zone est dédiée aux fiches des restaurants. Il ne peut figurer dans cette zone qu'une seule fiche à la fois. Ainsi il est possible de connaître le lieu, la spécialité, la gamme de prix et éventuellement l'ambiance du restaurant. Il y a également un plan de la situation géographique du restaurant avec les rues du quartier, les stations de métro, les monuments proches. Il est également possible de zoomer sur la carte et de se déplacer à l'intérieur. L'interaction avec la carte n'est cependant possible qu'avec la souris.

Les requêtes peuvent être formulées à l'oral (en cliquant sur le bouton "*Cliquez ici avant de parler*") ou à l'écrit (en tapant sa requête dans le champ de saisie) ou bien encore en cliquant sur des liens présents à l'écran (pour préciser un lieu, une spécialité ou encore consulter les solutions...). Il est possible de formuler sa requête en langage naturel à l'oral comme à l'écrit, mais également en utilisant seulement des mots clés. L'utilisateur peut donner ses trois critères dans un seul énoncé ou seulement un ou deux critère(s) et se laisser guider par le service pour affiner sa recherche. Il est possible de couper la synthèse vocale soit en prenant la parole, soit en validant une saisie écrite, ou encore en cliquant sur un lien. Si aucun restaurant ne répond aux critères de l'utilisateur, le service propose d'autres solutions dans les arrondissements voisins ou propose d'autres spécialités dans l'arrondissement demandé.

Deux phases décrivent une recherche typique avec Plan Resto. (1) Une phase d'émission des critères : l'utilisateur formule une demande plus ou moins précise en langage naturel ou en mot clés et le système l'aide en lui demandant de préciser ses critères, néanmoins, l'utilisateur peut, s'il le désire, fournir plusieurs critères en un tour de parole. (2) Une phase d'affinage : l'utilisateur a émis ses critères et suivant les réponses du système, il consulte et parcourt les solutions. Il peut aussi demander des informations ou des précisions sur les solutions. A tout moment, il peut basculer entre les phases 1 et 2.

Deux versions multimodales ont été utilisées pour les besoins de l'étude : une avec un avatar (figure 2) et une sans avatar (figure 3).

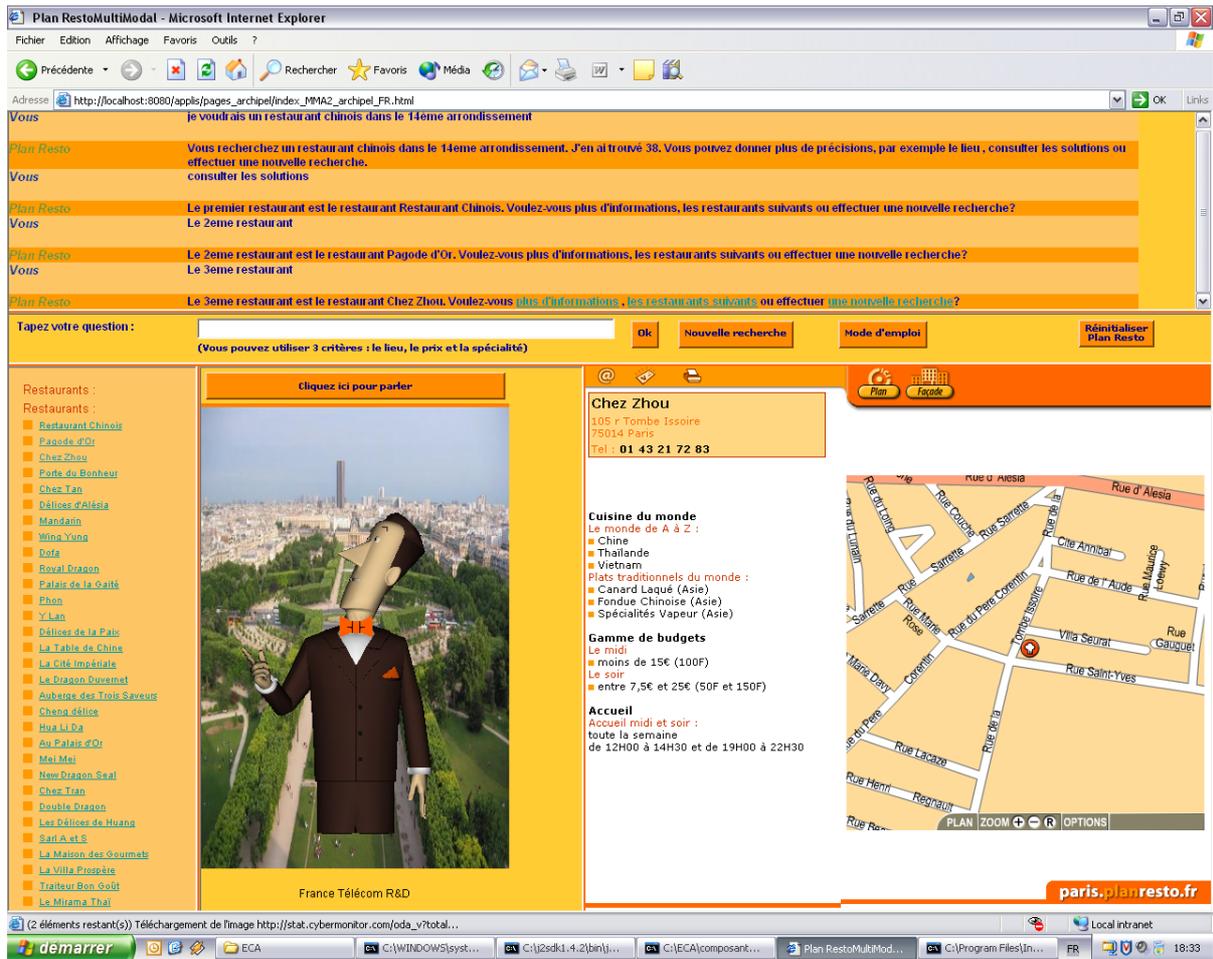


Figure 2 : Version multimodale avec avatar

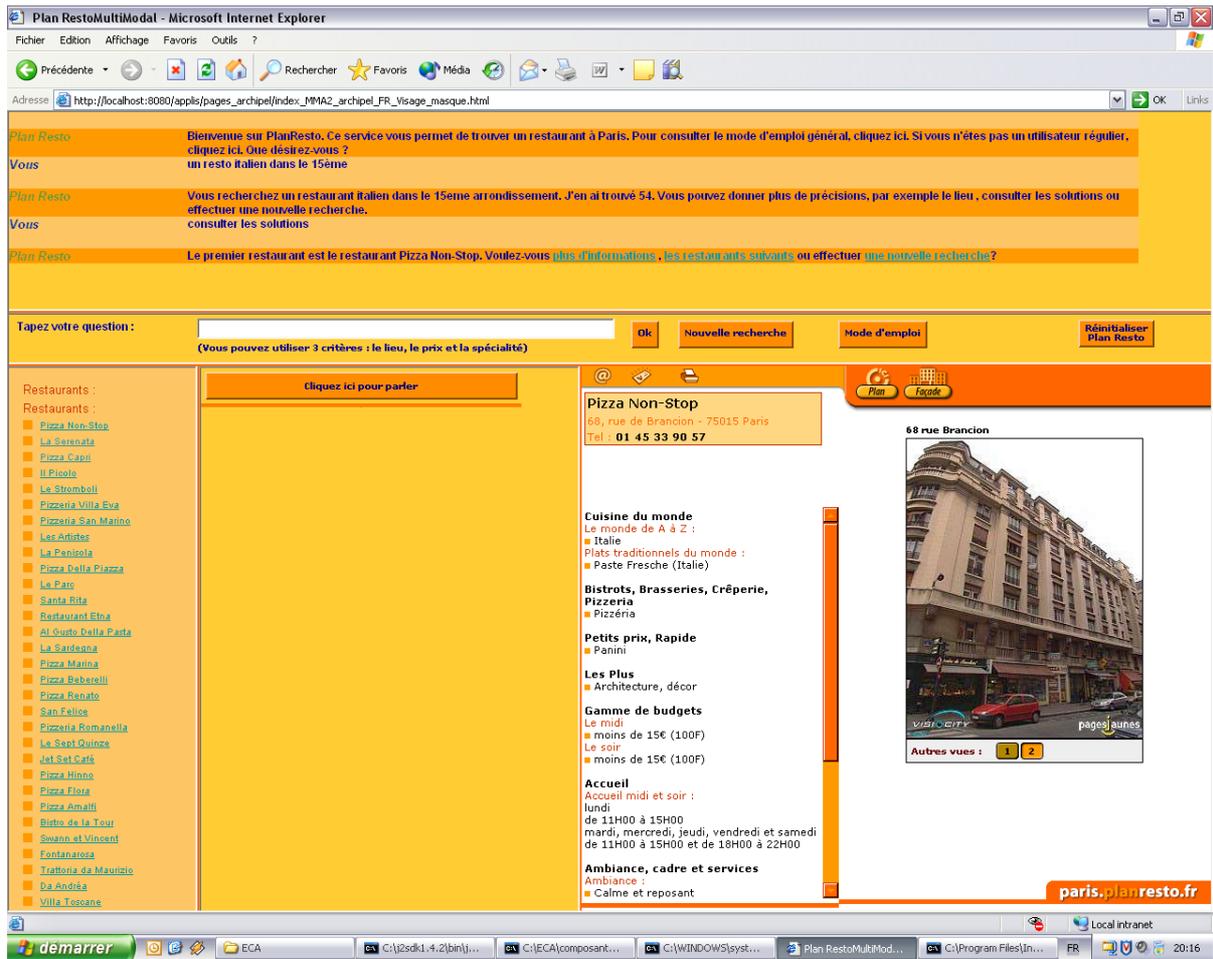


Figure 3 : Version multimodale sans avatar

### 3.3.2 Version graphique

Elle est identique à la version multimodale à la différence qu'il n'est pas possible de formuler sa requête à l'oral (il n'y a d'ailleurs pas de bouton dédié à cette fonction, ni de micro présent à côté du clavier) et il n'y a aucun son en sortie. De plus, il n'y a pas non plus de bouton "Cliquez ici pour parler", ni d'avatar (cf. figure 4).

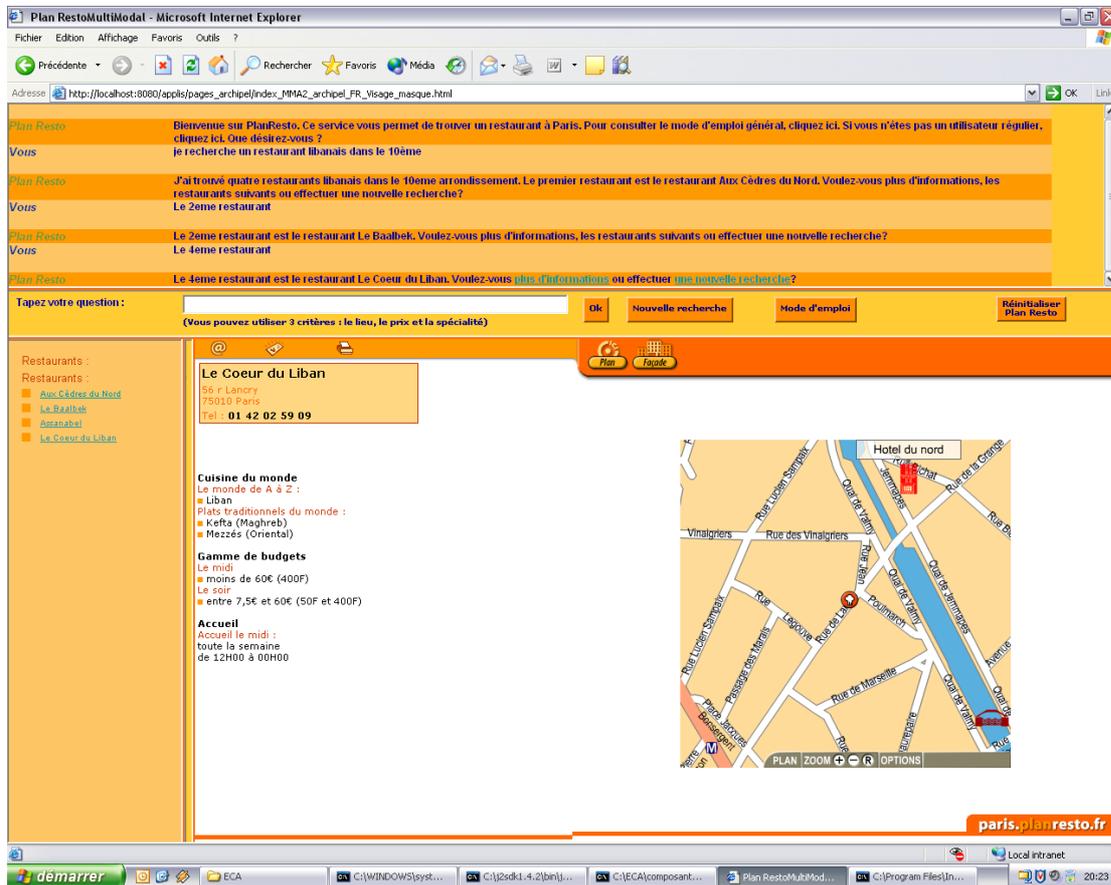


Figure 4 : Version graphique

### 3.3.3 Version vocale

Cette version ne fonctionne qu'à l'oral sur un téléphone fixe. Les énoncés vocaux sont les mêmes que ceux présents sur les versions graphique et multimodales. Le média utilisé (le téléphone) ne permet évidemment pas d'accéder au plan disponible dans les autres versions et il n'y a pas de description orale de l'emplacement du restaurant, hormis l'adresse (figure 5).

**Plan Resto** : Bienvenue sur le service Plan Resto [...]. Que désirez-vous ?  
**Utilisateur** : je recherche un restaurant dans le onzième arrondissement de Paris  
**PR** : Vous recherchez un restaurant dans le 11<sup>ème</sup> arrondissement. J'en ai trouvé plus de 10. Vous pouvez donner plus de précisions [...]  
**U** : Hum... un restaurant indien  
**PR** : Vous recherchez un restaurant indien dans le 11<sup>ème</sup> arrondissement. J'en ai trouvé 7. Vous pouvez donner plus de précisions, par exemple le lieu, consulter les solutions ou effectuer une nouvelle recherche.  
**U** : un restaurant pour vingt-cinq euros  
**PR** : Vous recherchez un restaurant indien pour environ vingt-cinq euros dans le 11<sup>ème</sup> arrondissement. J'en ai trouvé un qui s'appelle [...]. Désirez-vous plus d'informations ou effectuer une nouvelle recherche ?  
**U** : est-ce que vous pouvez répéter le nom du restaurant  
**PR** : [...]

Figure 5 : Exemple de dialogue oral avec Plan resto

### 3.3.4 Enregistrement vidéo

Une caméra numérique, installée sur un pied et positionnée au-dessus de l'épaule de l'utilisateur a permis un enregistrement audio-visuel du comportement du sujet et de ses réactions tout en conservant son anonymat, le visage n'étant pas filmé.

### 3.3.5 Questionnaires et entretiens

#### *Questionnaire de charge mentale*

La charge mentale a été évaluée à l'aide d'une version simplifiée et modifiée du test NASA-TLX (Task Load-Index, Hart et Staveland, 1988 in Le Bigot, 2004). Ce questionnaire est une échelle multidimensionnelle subjective bipolaire pondérée en 9 points, ayant pour but d'évaluer le niveau de charge de travail mentale imposée pour

une tâche particulière (cf. annexe 1). Il est composé de 6 questions faisant référence à 6 dimensions : (1) l'activité mentale (effort mentale), (2) la pression temporelle (enchaînement des sous-tâches), (3) l'effort (travail mental et physique), (4) la performance (efficacité de réalisation de la tâche), (5) la frustration, (6) l'activité physique (demande physique de réalisation de la tâche).

### *Questionnaire de satisfaction*

Un questionnaire de satisfaction, composé de 14 questions, a été élaboré à partir de 12 questions d'une étude antérieure réalisée sur Plan Resto (Le Bigot, 2004). Les réponses étaient données sur une échelle en cinq points (de "pas du tout d'accord" à "tout à fait d'accord"). Nous avons retenus les 14 questions pour l'analyse des résultats ( $\alpha = .88$ ) (cf. annexe 2).

### *Entretiens*

Une grille d'entretien concernant la multimodalité a été construite à partir de résultats d'études déjà réalisées sur des interfaces multimodales par France Télécom R&D. Cette grille a permis de réaliser des entretiens semi-directif à l'issue des tests pour les participants ayant utilisé une interface multimodale (cf. annexe 3). Les utilisateurs des versions monomodales (vocale et graphique) ont également participé à un "entretien libre" où ils pouvaient faire part de leur ressenti. (cf. annexe 4)

## **3.4 Procédure**

---

Les tests se sont déroulés à France Télécom R&D sur le site d'Issy-les-Moulineaux du 25 juillet au 4 août 2005. La séance débutait par l'explication des consignes, de l'objectif général des tests, de la façon dont les utilisateurs pourraient interagir avec le service "Plan Resto" et du déroulement des tests (cf. annexe 5). Les participants devaient ensuite réaliser successivement quatre scénarios de recherche de restaurant. Chaque participant n'utilisait qu'une version du service Plan resto (parmi les quatre versions existantes : vocale, graphique, multimodale sans avatar et multimodale avec avatar).

A la fin de chaque scénario, les participants devaient remplir un questionnaire de charge mentale. Puis à la fin des quatre scénarios ils remplissaient un questionnaire de satisfaction. Enfin, ils participaient à un entretien où ils pouvaient faire part de leur ressenti par rapport à l'interface et par rapport à la multimodalité (pour les utilisateurs d'interfaces multimodales). Les sessions de test duraient environ 45 minutes.

### 3.5 Scénarii

---

Les participants devaient trouver un restaurant à leur convenance à partir de quatre scénarii de recherche (cf. annexe 6).

Le **premier scénario** était une recherche libre, les utilisateurs devaient rechercher un restaurant à partir de leurs propres critères de recherche et ils n'avaient pour seule contrainte que de garder ces critères tout au long de leur recherche, autrement dit, il leur était demandé de ne pas changer d'avis en cours de recherche.

Le **deuxième scénario** était une recherche relativement simple dans la mesure où il leur était spécifié le lieu et la spécialité du restaurant à trouver et qu'il n'y avait qu'une seule réponse possible.

Le **troisième scénario** était plus difficile, les utilisateurs devaient trouver un restaurant dont la spécialité et l'arrondissement étaient spécifiés. Néanmoins, il se trouvait que l'arrondissement en question ne proposait pas la spécialité voulue. C'est pourquoi les utilisateurs devaient alors rechercher dans les arrondissements voisins, un restaurant à leur convenance. En outre, une inférence devait être faite par l'utilisateur sur l'intention d'être "*prêt à marcher*".

Enfin, le **quatrième scénario** a été construit dans le but d'observer comment les utilisateurs interagissaient lorsque la reconnaissance vocale ne fonctionne pas pour un lieu particulier. Ainsi, pour ce scénario, seule l'utilisation de la modalité graphique permettait d'outrepasser cette contrainte technique. C'est pourquoi, un autre lieu a été proposé aux utilisateurs de la version vocale, sans quoi il aurait été impossible pour eux de trouver un restaurant qui répondait au critère de lieu non reconnu par la reconnaissance vocale. Ce scénario ne sera pas pris en compte dans l'analyse des résultats d'un point de vue quantitatif, dans la mesure où tous les groupes n'ont pas eu la même recherche à effectuer pour ce scénario. Nous le retiendrons seulement pour l'analyse, en termes de fréquence d'utilisation des modalités, pour les groupes multimodale sans avatar et multimodale avec avatar.

### 3.6 Recueil des données

---

Les données ont été extraites des enregistrements vidéo, des réponses des utilisateurs aux questionnaires et des prises de notes effectuées lors des passations et lors des entretiens. Nous en avons retenus les indicateurs suivants (en gras dans le texte) :

### 3.6.1 En termes d'efficience

"L'efficience mesure l'économie en termes de temps de l'utilisation d'un objet finalisé et de l'atteinte du but recherché" (Tricot, 2000). Ainsi, nous relèverons comme indicateur de l'efficience :

- **Le temps** mis en secondes, pour effectuer chaque scénario.

### 3.6.2 En termes d'efficacité

"L'efficacité mesure l'économie en termes généraux (e.g. économie d'énergie, d'efforts, d'argent, de nombre de gestes ou de mouvements, d'attention, etc.) de l'utilisation d'un objet finalisé et de l'atteint du but recherché." (Tricot, 2000). Nous retiendrons comme indicateurs de l'efficacité :

- **Le nombre d'interactions** effectué pour réaliser chaque scénario de recherche : tours de parole, saisies au clavier et les clics sur les liens et les boutons. Cependant, nous n'avons pas pris en compte les clics effectués sur le plan, étant donné que cette interaction n'était pas possible pour la version téléphonique.
- **L'indice de charge mentale** pour chaque scénario (calculé à partir de la moyenne des six dimensions du questionnaire d'évaluation subjective de la charge de travail mentale).

### 3.6.3 En termes d'utilisation des modalités

- **La fréquence d'utilisation des trois modalités** possibles (parole, clavier, souris) pour chaque scénario (pour les utilisateurs des interfaces multimodales uniquement)

### 3.6.4 En termes de satisfaction

- **L'indice de satisfaction** pour chaque participant (calculé à partir de la moyenne des pondérations aux 14 questions de satisfaction)

De plus, nous avons pris en compte certaines données qualitatives : les verbalisations des sujets pendant les tâches de recherche et pendant les entretiens. Des notes ont également été prises au cours des passations concernant les interactions des participants avec Plan Resto.

### **3.7 Hypothèses**

---

Nous attendons un effet du facteur Version (interfaces multimodales vs. interfaces monomodales) :

- sur le temps mis pour effectuer les scénarii (inférieur pour les interfaces multimodales)
- sur le nombre d'interactions (inférieur pour les interfaces multimodales)
- sur la charge mentale (inférieure pour les interfaces multimodales)
- sur la satisfaction (supérieure pour les interfaces multimodales)

## 4. Résultats

### 4.1 Données quantitatives

#### 4.1.1 Temps mis pour effectuer chaque scénario

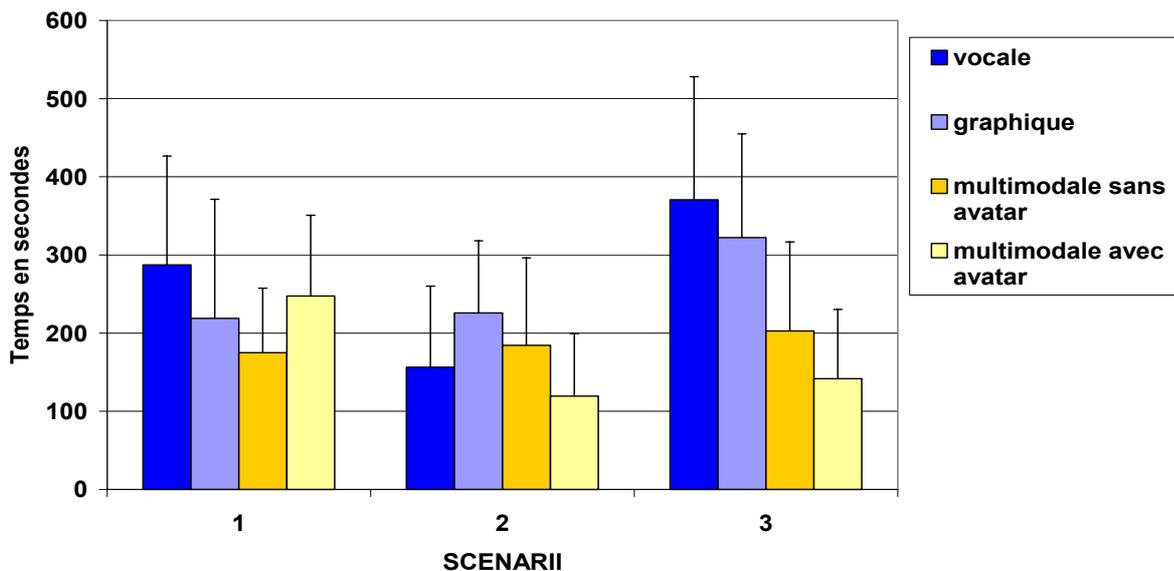


Figure 6 : Temps mis, en moyenne, par les quatre groupes de sujets (vocale, graphique, multimodale sans avatar et multimodale avec avatar) pour chaque scénario de recherche (les 3 premiers).

Une analyse de la variance réalisée sur l'ensemble des trois scénarii montre qu'il n'y pas d'effet significatif du facteur Version sur le temps mis pour effectuer l'ensemble des scénarii.

On observe néanmoins une tendance concernant l'effet de ce facteur ( $F(3,9) = 1,88$  ;  $p < .08$ ) (cf. annexe 7). En effet, les participants ayant eu des interfaces multimodales pour réaliser leurs recherches ont tendance à mettre moins de temps (170sec en moyenne pour la version avec avatar et 187sec en moyenne pour la version sans avatar) que les participants ayant eu des interfaces monomodales pour effectuer leurs recherches (256sec en moyenne pour la version graphique et 271sec en moyenne pour la version vocale).

De plus, une analyse de la variance réalisée sur chaque scénario pris isolément montre un effet du facteur Version pour le scénario 3 uniquement ( $F(3,20) = 4,22$  ;

$p < .05$ ) (cf. figure 6). Les comparaisons partielles effectuées sur cette analyse nous indiquent les différences significatives suivantes (cf. annexe 7) :

- Monomodales (vocale + graphique) vs. Multimodales (sans avatar + avec avatar) ( $F(1,20) = 11,51$  ;  $p < .01$ ) : les sujets des conditions monomodales mettent significativement plus de temps pour effectuer le scénario 3 que les sujets des conditions multimodales.

- Vocale vs. Multimodale sans avatar ( $F(1,20) = 5,33$  ;  $p < .05$ ) : les sujets du groupe Vocale mettent significativement plus de temps que les sujets du groupe Multimodale sans avatar.

- Vocale vs. Multimodale avec avatar ( $F(1,20) = 9,94$  ;  $p < .01$ ) : les sujets du groupe Vocale mettent significativement plus de temps que les sujets du groupe Multimodale avec avatar.

- Graphique vs. Multimodale avec avatar ( $F(1,20) = 6,19$  ;  $p < .05$ ) : les sujets du groupe Graphique mettent significativement plus de temps que les sujets du groupe Multimodale avec avatar.

Pour résumer, nous observons une tendance pour les interfaces multimodales à être plus efficaces que les interfaces monomodales, en termes de temps mis pour effectuer les scénarios, surtout quand la tâche est complexe (scénario 3).

#### 4.1.2 Nombre d'interactions effectué pour chaque scénario

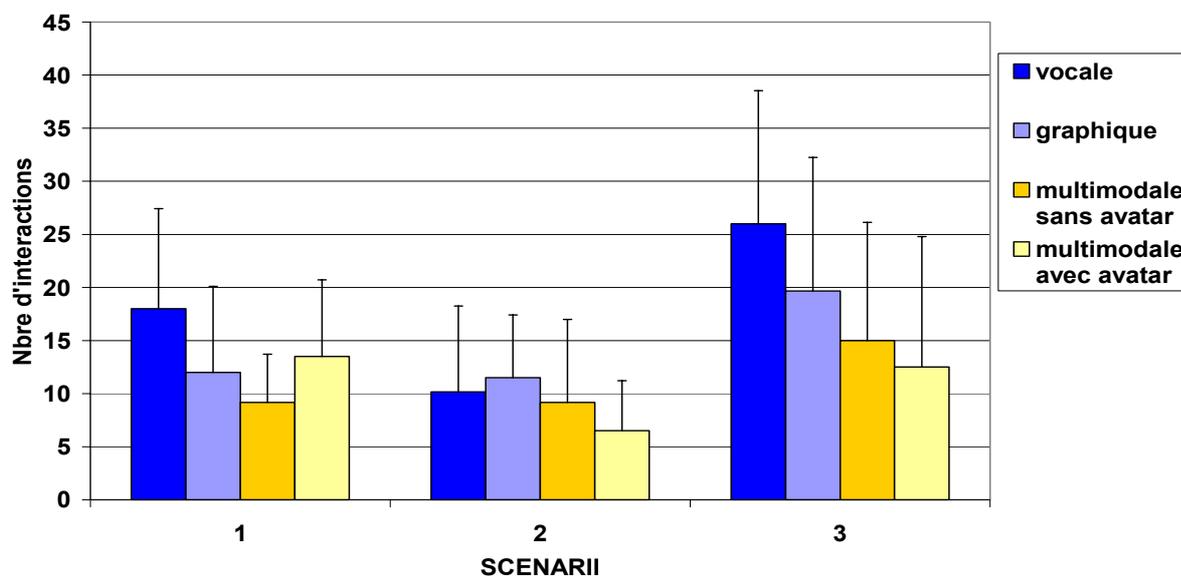


Figure 7 : Nombre d'interactions effectué, en moyenne, par les quatre groupes de sujets (vocale, graphique, multimodale sans avatar et multimodale avec avatar) pour chaque scénario de recherche (les 3 premiers).

Une analyse de la variance réalisée sur l'ensemble des trois scénarii montre qu'il n'y pas d'effet significatif du facteur Version sur le nombre d'interactions effectué pour l'ensemble des scénarii (cf. annexe 8).

En outre, une analyse de la variance réalisée sur chaque scénario pris isolément ne montre aucune différence significative concernant le nombre d'interactions effectué entre les différentes conditions (cf. annexe 8).

Cependant, nous pouvons observer que les sujets des conditions monomodales auraient tendance, pour le scénario 3 (le plus complexe), à effectuer plus d'interactions que les sujets des conditions multimodales (cf. figure 7).

#### 4.1.3 Fréquence d'utilisation des différentes modalités (pour les interfaces multimodales uniquement)

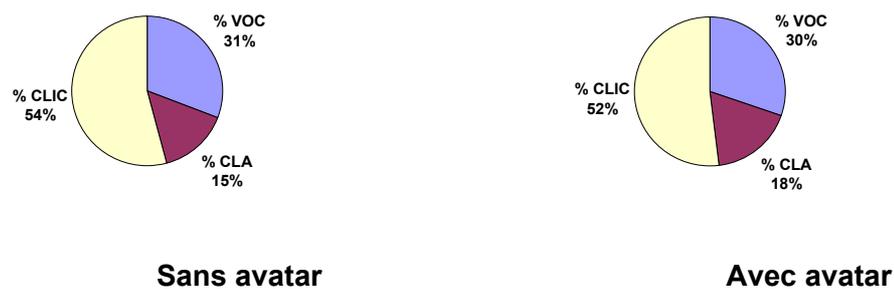


Figure 8 : Fréquence d'utilisation des différentes modalités (CLIC = clics souris ; CLA = saisies au clavier ; VOC = utilisation de la parole) sur l'ensemble des scénarii pour les conditions multimodales sans avatar et avec avatar.

La figure 8 ci-dessus montre qu'il n'y a pas de différence entre les sujets de la condition multimodale avec avatar et multimodale sans avatar, en termes de modalités utilisées, en moyenne par groupe, pour l'ensemble des scénarii.

Cependant, on constate que les 2 modalités les plus "rapides" pour l'interaction en entrée sont les plus utilisées : les clics et la modalité vocale.

#### 4.1.4 Profils d'utilisation des modalités (pour les interfaces multimodales uniquement)

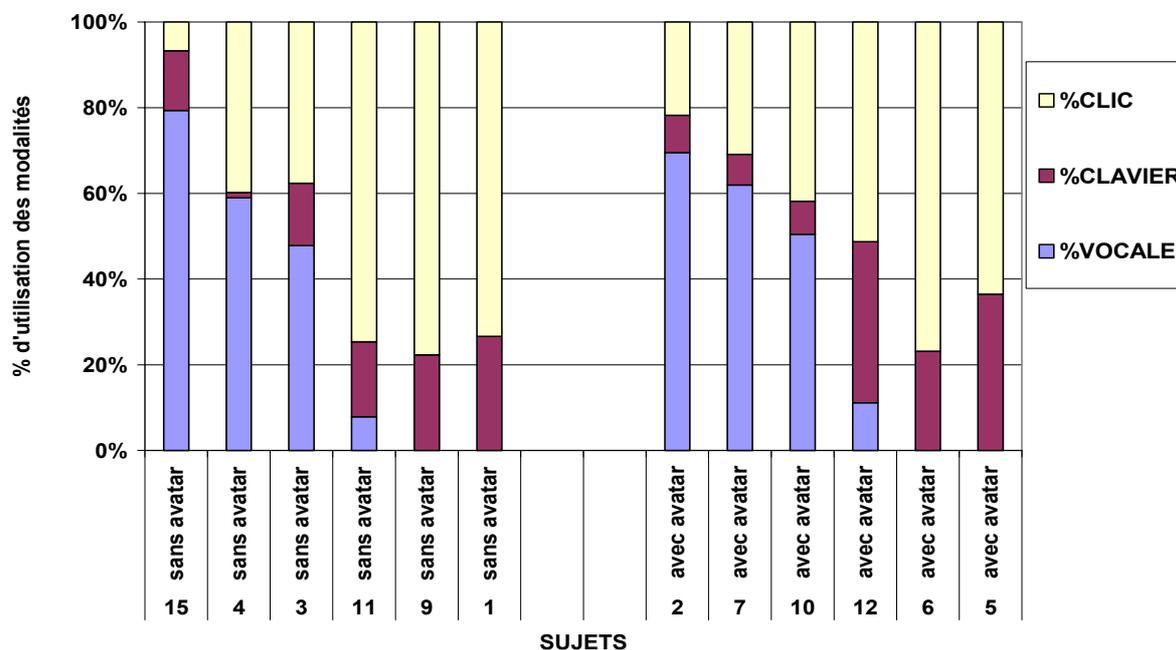


Figure 9 : Fréquence d'utilisation des modalités pour les six sujets de la condition multimodale sans avatar (à gauche) et pour les six sujets de la condition multimodale avec avatar (à droite) sur l'ensemble des scénarii. Chaque sujet est désigné par son numéro d'ordre de passation.

Cet histogramme (figure 9) permet d'observer que l'on retrouve des profils d'utilisation des modalités relativement similaires entre les deux conditions (multimodale sans avatar et multimodale avec avatar) concernant l'utilisation des trois modalités qui étaient à leur disposition (clic, clavier et vocale) :

- des utilisateurs qui ont utilisé beaucoup le vocal, peu le clavier et peu de clics (sujet 15 sans avatar et sujet 2 avec avatar)
- des utilisateurs qui ont utilisé moyennement le vocal, peu le clavier et beaucoup de clics (sujets 4 et 3 sans avatar et sujets 7 et 10 avec avatar)
- des utilisateurs qui ont peu utilisé le vocal, un peu plus le clavier et beaucoup de clics (sujet 11 sans avatar et sujet 12 avec avatar)
- des utilisateurs qui n'ont **jamais** utilisé la modalité vocale, un peu le clavier et beaucoup de clics (sujets 9 et 1 sans avatar et sujets 6 et 5 avec avatar)

Ainsi, la présence de l'avatar ne semble pas influencer le comportement des participants concernant les modalités qu'ils utilisent. Au contraire, on retrouve pratiquement les mêmes profils d'utilisation dans les deux groupes (avec et sans avatar).

Pour résumer, nous observons qu'il existe différents profils d'utilisation des modalités pour les interfaces multimodales. De plus, on retrouve ces différents profils dans des proportions similaires entre la version sans avatar et la version avec avatar.

#### 4.1.5 Indice de charge mentale

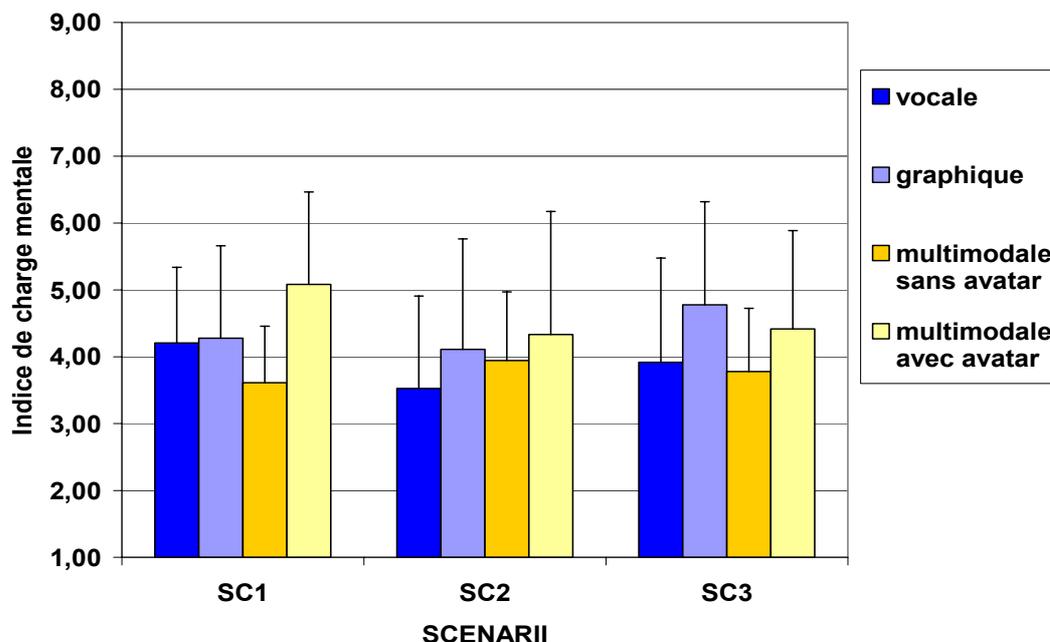


Figure 10 : Indice de charge mentale, en moyenne, pour les quatre groupes de sujets (vocale, graphique, multimodale sans avatar et multimodale avec avatar) pour chaque scénario de recherche (les 3 premiers).

Une analyse de la variance réalisée sur l'ensemble des trois scénarii montre qu'il n'y pas d'effet significatif du facteur Version sur l'indice de charge mentale (cf. annexe 9). En outre, une analyse de la variance réalisée sur chaque scénario pris isolément ne montre aucune différence significative concernant l'indice de charge mentale entre les différentes conditions (cf. annexe 9).

Cependant, la figure 10 permet d'observer des tendances montrant que l'appréciation subjective des participants sur leurs charges mentales serait plus importante en utilisant la version multimodale avec avatar qu'en utilisant la version multimodale sans avatar, surtout pour le scénario 1 qui constitue en lui-même la "prise en main" de l'interface. De plus, la version graphique montrent également une tendance à un ressenti de plus forte charge mentale qu'avec la version vocale. On peut remarquer que l'appréciation subjective de la charge mentale est supérieure pour l'interface graphique par rapport à la version multimodale sans avatar.

### 4.1.6 Indice de satisfaction

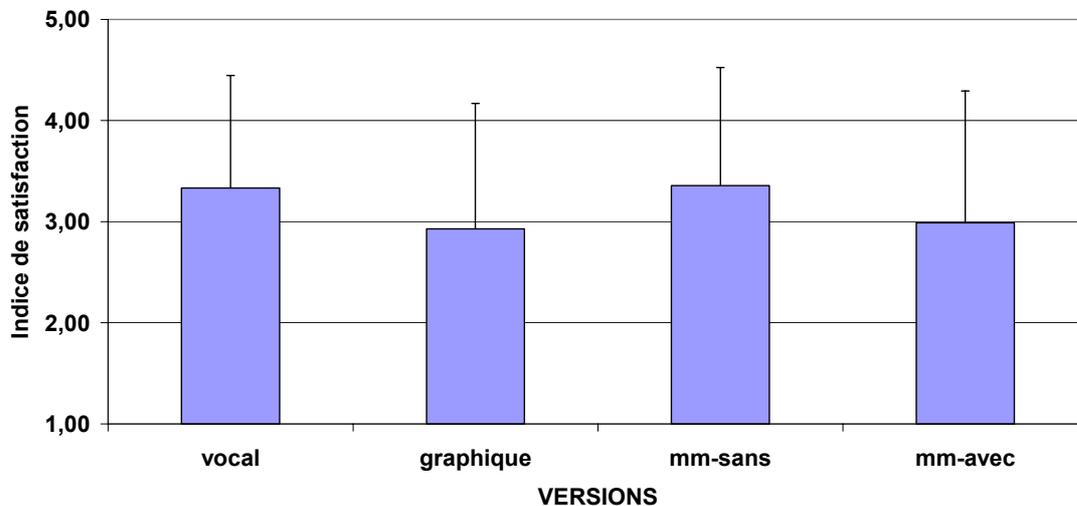


Figure 11 : Indice de satisfaction, en moyenne, pour les quatre groupes de sujets (vocale, graphique, multimodale sans avatar et multimodale avec avatar) sur leurs versions de Plan Resto.

Une analyse de la variance (cf. annexe 10) montre qu'il n'y a pas d'effet significatif du facteur Version sur l'indice de satisfaction des participants.

Néanmoins, la figure 11 semble montrer que ce sont les versions multimodale sans avatar et vocale, qui obtiennent des scores de satisfaction les plus élevés.

## 4.2 Données qualitatives

### 4.2.1 Concernant la multimodalité en entrée

*La modalité vocale :*

Nous avons pu remarquer que plus de 87% des participants (7 utilisateurs sur 8 ayant utilisé la modalité vocale) ont oublié plusieurs fois de cliquer sur le bouton "cliquez ici pour parler" avant de parler. Ceci peut s'expliquer par le fait que certains utilisateurs avaient "[...] envie de répondre spontanément et pas d'appuyer sur un bouton avant de parler" (sujet 5) ou encore à cause du bouton "cliquez ici pour parler" qui n'est "pas assez saillant" (sujets 3, 9 et 10). De plus, la reconnaissance vocale, une fois activée (appui sur le bouton), ne fonctionne que pendant quelques

secondes. Ainsi, nous avons pu observer que lorsque les utilisateurs étaient hésitants ou lorsqu'ils prenaient leur temps pour parler, la reconnaissance vocale ne prenait pas en compte l'ensemble de leur requête.

Par ailleurs, certains participants n'ont jamais utilisé la modalité vocale. Quand nous leur avons demandé pourquoi, ils ont expliqué ce choix pour les raisons suivantes : soit *"par habitude sur Internet"* (sujet 1), soit parce que *"ça va plus vite avec la souris"*(sujet 4).

#### *Le clavier et la souris :*

Certains utilisateurs ont trouvé que taper les requêtes était un moyen *"plus sûr"*(sujet 2), *"plus facile"* (sujet 3) ou encore plus rapide : *"ça va plus vite à l'écrit"* (sujet 11). En outre, cette modalité a été souvent utilisée lorsque la reconnaissance vocale faisait trop d'erreurs.

La modalité "souris" a été très utilisée par les participants, ceci se comprend aisément dans la mesure où cette modalité est la plus rapide : *"je préfère utiliser les liens plutôt que le micro, ça va plus vite"* (sujet 12)

#### *Le choix et le passage d'une modalité à l'autre :*

Certains utilisateurs ont trouvé "[...] *rassurant qu'il y ait les deux [vocale et clavier]*" (sujet 3), et d'autres ont même ajouté qu'*"heureusement, on peut changer de modalité"* (sujets 4 et 7). Enfin certains pensent que *"c'est bien de pouvoir utiliser différentes modalités mais il faut qu'elles soient équivalentes au niveau de la performance"* (sujet 10).

Souvent le passage d'une modalité à l'autre est motivé par la volonté de pouvoir outrepasser les erreurs du système : *"quand j'ai échoué à l'oral pour [la station de métro] Filles du Calvaire, je suis passé à l'écrit"* (sujet 11) et inversement le sujet 3 à écrit "30" pour 30 euros, le système ne l'a pas pris en compte, puis elle a dit *"pour 30 euros"* à l'oral et elle a ainsi pu continuer sa recherche.

## **4.2.2 Concernant la multimodalité en sortie**

#### *La redondance écrit / synthèse vocale :*

Plus de 66% (8 sujets sur 12) des sujets ont trouvé la redondance gênante, certains ont même essayé de couper les haut-parleurs à l'aide de l'icône présent dans la barre des tâches. De même, plus de 66% des participants (8 sujets sur 12) ont émis le souhait de pouvoir couper la voix ou d'"*au moins pouvoir faire une pause*" (sujet 3).

Face à cette redondance, les participants semblent choisir entre deux modes d'interaction : soit ils essaient de se "couper" du vocal "*quand j'ai vu qu'il disait la même chose, j'ai arrêté de l'écouter*" (sujets 6), certains allant même jusqu'à adopter des stratégies pour stopper la synthèse vocale "*heureusement qu'il y a des liens pour le stopper*" (sujet 7) ; soit au contraire ils se focalisent sur le dialogue à l'oral "*j'ai découvert l'historique [du dialogue avec Plan Resto] seulement au bout du 4<sup>ème</sup> scénario, je me suis plutôt focalisé sur le champ blanc et sur l'interaction vocale*" (sujet 15).

Enfin, un certain nombre de sujets préféreraient avoir une complémentarité entre ce qui est dit et les informations à l'écran, plutôt que de la redondance totale.

#### *L'historique du dialogue avec Plan Resto :*

Les participants ont perçu l'historique surtout comme un bon moyen de savoir ce que le système avait compris de ce qu'ils disaient. En d'autres termes, l'historique constitue un bon feedback pour les utilisateurs : "*l'intérêt du textuel [l'historique] réside surtout dans le fait qu'on peut savoir ce qu'il a compris de ce que l'on dit*" (sujet 7) ; "*c'est pratique de savoir ce qu'il a compris*" (sujet 2).

#### *La voix :*

Presque la moitié des sujets (5 sujets sur 12) a trouvé la voix désagréable : "*la voix est abominable*" (sujet 3) ; "*sa façon de parler est énervante*" (sujet 5) ; "*la voix fait peur, il est un peu menaçant*" (sujet 9). La vitesse d'élocution de la synthèse vocale a également gêné certains utilisateurs : "*il parle trop doucement*" (sujet 6).

#### *L'avatar :*

Certains participants déclarent ne pas s'être focalisé sur l'avatar : "*je n'ai pas regardé l'avatar*" (sujet 6) ; "*on remarque l'avatar, mais on est plus sensible à l'information textuelle*" (sujet 5).

En outre, d'autres participants n'ont pas apprécié l'avatar : "*l'avatar est très énervant*" (sujet 7), "*l'avatar ne met pas à l'aise, on a peur de lui couper la parole*" (sujet 5) ; "*il faudrait pouvoir supprimer l'avatar*" (sujet 12) ou "*le désactiver*" (sujet 5) ; "*l'avatar prend une place énorme alors qu'il n'est pas nécessaire, il ne sert à rien, il est inefficace*" (sujet 10).

## **4.3 Discussion des résultats**

---

### **4.3.1 Concernant l'efficience**

Concernant le temps mis pour effectuer les scénarios, les interfaces multimodales n'ont été significativement plus efficaces par rapport aux interfaces monomodales, que pour le scénario 3. Ainsi notre hypothèse concernant l'efficience n'est que partiellement validée, à savoir : les interfaces multimodales sont plus efficaces que les interfaces monomodales (en termes de temps mis pour réaliser les scénarii de recherche). Cependant, il est intéressant de noter que le scénario 3 constitue la recherche la plus difficile à effectuer.

En effet, il est demandé aux participants de chercher un restaurant végétarien dans le 15<sup>ème</sup> arrondissement alors que la base de données n'en contient pas. Le système propose alors des restaurants dans les arrondissements proches. La consultation des différentes solutions s'est alors trouvée facilitée par la possibilité d'avoir plusieurs modalités à sa disposition pour en trouver une. Autrement dit, lorsque le système se met à proposer des alternatives, la multimodalité serait plus efficace, tandis que lorsque les utilisateurs savent exactement ce qu'ils veulent (scénario 1) ou lorsqu'il n'y a qu'une solution possible (scénario 2) la multimodalité ne permettrait pas d'être significativement plus efficace en termes de temps. Ceci est d'ailleurs en lien avec certains propos des utilisateurs d'interfaces monomodales (vocale et graphique) : "*Plan Resto c'est bien quand on sait exactement ce qu'on veut, mais sinon c'est pas vraiment évident...*"(sujet 16).

### **4.3.2 Concernant l'efficacité**

Bien que nous n'ayons pas trouvé de résultat statistiquement significatif concernant notre hypothèse sur l'efficacité en termes de nombre d'interactions, nous pouvons cependant noter une certaine tendance des interfaces multimodales à être plus efficaces que les interfaces monomodales en termes de nombre d'interactions pour

réaliser les recherches concernant le scénario 3 qui constitue le scénario le plus complexe des tests.

Néanmoins, nous avons pu observer lors des tests que la multimodalité avait bien servi à lever l'ambiguïté générée par la synthèse vocale. En effet, dans la version vocale, des sujets n'ont pas compris que le système leur posait une question. Les participants devaient faire un choix entre deux propositions pour poursuivre leurs recherches et ils ont compris que le système faisait une affirmation en prenant en compte les deux propositions du choix ("*vous recherchez un restaurant végétarien dans le 15<sup>ème</sup> arrondissement ou dans le 16<sup>ème</sup> arrondissement ?*"). A l'inverse, aucun sujet des conditions multimodales n'a pris cet énoncé pour une affirmation, car il y avait la présence du point d'interrogation dans la proposition écrite.

L'hypothèse concernant l'efficacité en termes d'indice de charge mentale plus faible pour les interfaces multimodales par rapport aux interfaces monomodales n'a pas été démontrée statistiquement. Notons quand même que l'appréciation subjective des utilisateurs montre que la version multimodale avec avatar semble demander plus de ressources cognitives que les interfaces multimodales sans avatar. De plus, les participants de la condition graphique, ont évalué plus lourdement la charge mentale pour le scénario 3 que les sujets de la condition multimodale sans avatar qui est pourtant visuellement très proche.

Cependant, l'estimation de la charge mentale à été effectuée dans cette étude à partir d'un questionnaire subjectif. Une méthodologie plus objective telle que le paradigme de la "double tâche" permettrait de vérifier et de préciser les résultats obtenus concernant la charge mentale des sujets lors de la réalisation des tâches. Cette technique consiste à faire accomplir à un sujet deux tâches en parallèle (une tâche principale et une tâche secondaire) et d'observer les variations de performance à la tâche principale ou à la tâche secondaire, selon le protocole choisi (Falzon, 2004). Néanmoins, ce paradigme s'avère beaucoup moins économique à mettre en œuvre qu'un questionnaire à compléter *a posteriori*. Enfin, le concept de charge mentale est particulièrement discuté en psychologie cognitive et en ergonomie quant à sa définition ou plutôt ses définitions relativement floues et dont "*la mesure [...] en termes absolus pose trop de problèmes pour être envisagée sérieusement*" (Tricot et Chanquoy, 1996).

### **4.3.3 Concernant les modalités utilisées**

Dans notre étude, on observe dans les interfaces multimodales, que la souris est beaucoup utilisée. L'aspect "rapidité d'interaction" est donc une donnée importante dans la conception des interfaces multimodales. Il serait intéressant à cet égard, d'observer les modalités utilisées sur un mobile qui n'aurait pas de moyen de sélection rapide de type "souris". Ainsi, il serait possible de tester si le vocal prend

l'avantage en termes de modalité utilisée et d'observer comment la modalité "clavier" évolue.

Par ailleurs, il est intéressant de noter que les interfaces multimodales ne sont pas utilisées de la même façon par tous les sujets. Ceci rejoint l'argument d'Oviatt (2003) quant à la propension des interfaces multimodales à permettre une certaine flexibilité. Dans une autre étude (Oviatt, 1999), elle avait également mis en évidence de fortes différences interindividuelles concernant l'intégration des commandes multimodales.

Enfin, on peut remarquer que l'avatar n'a pas eu d'effet sur les modalités utilisées et ceci se retrouve dans les verbalisations post-test des sujets. En effet, l'avatar semble avoir été intentionnellement ignoré soit par agacement, ou bien encore parce qu'il était tout simplement possible de pouvoir réaliser ses recherches sans lui.

#### **4.3.4 Concernant la redondance**

La redondance des informations écrites et sonores (synthèse vocale) en sortie semble avoir particulièrement gêné les utilisateurs. Moreno et Mayer ont démontré dans une étude où ils concluent que *"le caractère interférent de la redondance n'apparaît qu'en situation de présentation simultanée, propice à l'apparition de surcharge en mémoire de travail"* (Moreno et Mayer, 2002 in Le Bohec et Jamet, 2005).

Dans notre étude, pour pallier cette surcharge d'informations, les participants semblent choisir un mode d'interaction particulier : soit plutôt centrés sur le "graphique", soit plutôt sur le "vocal". Ainsi, ils ne changent de modalité que lorsque celle qu'ils utilisent n'est plus performante à un moment donné. Ceci peut expliquer, en partie, les différences de profils d'utilisation des modalités.

#### **4.3.5 Concernant la satisfaction**

Notre hypothèse concernant la satisfaction n'a pas été statistiquement validée. Néanmoins les résultats semblent donner un petit avantage à la version multimodale sans avatar, comparée à la version multimodale avec avatar.

En outre, les données qualitatives semblent indiquer que la perception du service est particulièrement affectée par la qualité, la vitesse d'élocution et la prosodie de la synthèse vocale. Par ailleurs, l'efficacité de la reconnaissance vocale semble jouer un rôle primordial sur la satisfaction des utilisateurs. Ainsi, lors des tests, on a pu voir

des utilisateurs abandonner la modalité vocale au profit de la modalité graphique, jugée plus fiable et sur laquelle l'utilisateur peut exercer un contrôle plus accru.

Enfin, bien que beaucoup de nos hypothèses n'aient pas été statistiquement validées, nous avons tout de même pu observer certaines tendances. De plus, nos quatre groupes expérimentaux ne sont composés que de six sujets chacun. Il n'est donc pas évident de pouvoir faire ressortir des différences statistiquement significatives avec si peu de sujets. Par ailleurs, les différentes versions du service Plan Resto ne sont actuellement qu'à l'état de maquette et elles n'ont pas été développées, pour l'instant, dans un souci totalement ergonomique.

Néanmoins, ces résultats nous ont permis d'élaborer des recommandations ergonomiques qui seront autant de pistes de réflexions pour la suite du projet VERBATIM.

## 5. Recommandations

### 5.1 Concernant la multimodalité en entrée

---

#### 5.1.1 Le choix entre modalités

**Reco.** *Avoir une équivalence des modalités en termes de performances\* et d'actions possibles. En effet, cela permettra à l'utilisateur de choisir la modalité avec laquelle il se sentira le plus à l'aise et cela permettra une utilisation plus aisée dans différents contextes, notamment en mobilité (forte luminosité, ambiance bruyante, etc.).*

*\* Il s'agit ici, autant que possible, de faire en sorte que les performances, en termes de temps de traitement de la requête par le système, soient à peu près équivalentes d'une modalité à l'autre.*

Concernant la notion de performance d'une modalité par rapport à une autre, nous parlons ici en termes de différence de temps de réaction du système et du caractère acceptable ou non de ce temps pour l'utilisateur. En effet, il semblerait que la flexibilité que peut induire une interface multimodale peut être contrebalancée par le fait que le système s'avère moins réactif dans une modalité par rapport à une autre.

Cette recommandation est fortement liée aux limitations techniques existantes encore à l'heure actuelle concernant les technologies de reconnaissance vocale.

Néanmoins, d'autres études [Karsenty06], montrent que les déterminants du choix d'une modalité n'est pas uniquement basé sur la performance du système mais aussi sur :

- la tâche,
- des préférences initiales qui prédisposeraient à opter ou à délaissé dès le départ certaines modalités,
- l'expérience de l'utilisateur qui semble pouvoir modifier en partie les préférences initiales des utilisateurs (développement de l'usage combiné de plusieurs modalités et assignation de certaines modalités pour certaines commandes).

Le second point va dans le sens de nos résultats sur l'usage des modalités (Cf. livrable Verbatim sur l'expérimentation d'un service multimodal). En effet, ces résultats montrent une forte variabilité inter-individuelle concernant la modalité vocale

presqu'aussi performante que la modalité graphique en termes de temps de réponse et parfois jamais utilisée.

Par ailleurs, la plupart des prototypes de services vocal-graphique que nous avons pu manipuler présentaient un net décalage de temps de réponse entre la modalité graphique (réponse immédiate) et la modalité vocale (quelques secondes d'attente). Plusieurs tests utilisateurs réalisés sur de tels prototypes ont montré une utilisation du vocal rapidement diminuée au profit du graphique.

### 5.1.2 La modalité vocale

**Reco.** *Prévoir un bouton push-to-talk (l'utilisateur maintient le bouton appuyé jusqu'à ce qu'il ait fini de parler) afin que l'utilisateur puisse contrôler le temps qui lui est imparti pour parler. Cette recommandation est d'ailleurs appuyée par d'autres études réalisées à France Télécom R&D.*

**Reco.** *Afficher de façon saillante le bouton ou l'icône qui permet de prendre la parole. En effet, certains utilisateurs ont trouvé ce bouton trop peu saillant.*

## 5.2 Concernant la multimodalité en sortie

---

### 5.2.1 Complémentarité modale et redondance

**Reco.** *Choisir un ou plusieurs types de coopérations entre les modalités de façon pertinente. En effet, la redondance ne saurait se justifier pour tous les énoncés en sortie du système, elle est à privilégier pour les messages d'erreurs et/ou pour les feedbacks informant que sans une modification de la requête de l'utilisateur, celle-ci ne pourra pas aboutir.*

**Reco.** *Favoriser la complémentarité modale plutôt que la redondance totale et réserver la redondance pour des messages d'alerte ou de haute importance par exemple.*

**Reco.** *Eviter la redondance totale pour les énoncés longs. En effet, cela risque d'entraîner une surcharge mentale pour les utilisateurs.*

## 5.2.2 La synthèse vocale

**Reco.** *Permettre à l'utilisateur de désactiver la synthèse vocale. Le système doit lui permettre d'avoir un contrôle permanent sur la synthèse vocale à l'aide d'une icône dédiée et d'une commande vocale (exemple : "couper le son !"). L'utilisateur doit également pouvoir faire une pause et reprendre à l'endroit où il avait arrêté la synthèse vocale. Il doit aussi pouvoir faire répéter le dernier énoncé de la synthèse vocale.*

**Reco.** *Permettre à l'utilisateur de personnaliser la voix de la synthèse vocale\* en lui laissant le choix entre plusieurs voix (féminines et masculines). En effet, le ton de la voix semble jouer un rôle important quant à la satisfaction des utilisateurs.*

*\* Cette recommandation est à la limite de l'ergonomie, elle vise à tenter de mieux faire accepter la synthèse vocale. Mais il ne faut pas que ce soit au détriment de l'utilisabilité générale du service, à savoir que cette fonctionnalité supplémentaire "choix de la voix" ne doit pas alourdir l'interface.*

**Reco.** *Permettre à l'utilisateur de contrôler la vitesse d'élocution de la synthèse vocale. En effet, les utilisateurs ont trouvé, en général, que la synthèse vocale était trop lente.*

## 5.2.3 Feedback

**Reco.** *Donner un feedback écrit aux utilisateurs de ce que la reconnaissance vocale a pris en compte. Ce feedback paraît en effet nécessaire d'autant plus que la modalité vocale est jugée moins sûre que la modalité graphique.*

## 5.2.4 L'avatar

**Reco.** *Laisser la possibilité à l'utilisateur de désactiver l'avatar. En effet, certains participants ont été gênés par sa présence.*

**Reco.** *Permettre à l'utilisateur de personnaliser son avatar pour une meilleure acceptation de ce dernier\*.*

*\* Cette dernière recommandation est à la limite de l'ergonomie, elle vise à tenter de mieux faire accepter l'avatar. Mais il ne faut pas que ce soit au détriment de l'utilisabilité générale du service, à savoir que cette fonctionnalité supplémentaire "choix de l'avatar" ne doit pas alourdir l'interface.*

## 6. Conclusion, perspectives

Cette étude ergonomique réalisée sur le service "Plan Resto" a permis d'élaborer des recommandations ergonomiques pour la multimodalité et de mettre en évidence certains aspects de l'interaction multimodale. En effet, nous avons observé une tendance pour ces interfaces à être plus efficaces et un peu plus efficaces que des interfaces monomodales, grâce à des mesures quantitatives. Ceci s'amplifie notablement lorsque la tâche demandée aux sujets est plus complexe. Les résultats qualitatifs ont quant à eux montré que : les interfaces multimodales ne sont ni rejetées, ni préférées, mais qu'en particulier, elles sont ressenties comme rassurantes en cas de difficultés (scénario complexe), car elles permettent à l'utilisateur de changer de moyen d'interaction. En revanche, elles demandent à être aménagées et précisément adaptées aux besoins des utilisateurs du service pour constituer réellement un apport (par exemple, pouvoir supprimer le vocal à la demande). Ainsi, un axe de recherche pourrait être d'introduire, de façon systématique et contrôlée, une variable "complexité de la tâche", afin de voir si les résultats de cette étude pourraient être confirmés et affinés.

De plus, notre étude a pu mettre en évidence que les interfaces multimodales ne sont pas utilisées de façon uniforme par tous les utilisateurs et que la présence de l'avatar n'affecte pas le "comportement modal" des sujets.

Il serait intéressant de compléter ces tests par une étude des mouvements oculaires afin de pouvoir observer objectivement les zones de l'écran parcourues à différents moments de l'interaction. Ainsi, il serait possible d'observer plus finement ce que les sujets regardent lorsqu'il y a une synthèse vocale et un avatar par exemple, et ce qui se passe lorsqu'il y a redondance en sortie (est ce que les sujets lisent en même temps ou non).

Un des enjeux, pour les prochaines études sur ce sujet, sera d'observer si l'on retrouve le même type de résultats sur un terminal mobile, concernant la multimodalité en entrée comme en sortie. En effet, la redondance totale n'est pas réellement utile pour Plan Resto sur PC (web). Cependant, est-ce également le cas sur un mobile, si l'on fait varier différentes contraintes environnementales comme des bruits extérieurs ou une forte luminosité ? Par ailleurs, ces prochains tests pourraient être réalisés sur un nombre de participants plus important afin d'obtenir des résultats statistiques plus robustes. Enfin, il pourrait être intéressant de comparer les ressentis des utilisateurs lors de tests ergonomiques où ils verraient différentes versions d'un même service.

## 7. Bibliographie

André, E., Rist, T., Muller, J (1998). Integrating reactive and scripted behaviors in a life-like presentation agent. Proceedings of AGENTS'98, pp. 261-268. May 9-13, Minneapolis/St. Paul.

Bellik, Y. (1995). *Interfaces multimodales : concepts, modèles et architectures* (Thèse de doctorat 30 mai 1995). Université de Paris-Sud Orsay.

Buisine, S. (2005). *Conception et évaluation d'agents conversationnels multimodaux bidirectionnels* (Thèse de doctorat 2005). Université de Paris V.

Falzon, P. & Sauvagnac, C. (2004). Charge de travail et stress. In P. Falzon (Ed.) *Ergonomie* (pp. 175-190). Paris : P.U.F.

Le Bohec, O., & Jamet, E. (2005). Les effets de la redondance dans l'apprentissage à partir de documents multimédia. *Le travail humain*, 68(2), 97-124.

Le Bigot, L. (2004). *La recherche d'informations avec un système de dialogue en langage naturel* (Thèse de doctorat décembre 2004). Université de Poitiers.

Nigay L. (2001) *Modalité d'interaction et multimodalité* Rapport en vue de l'habilitation à diriger des recherches. Université de Grenoble.

Oviatt, S.L. (1999). *Ten myths of multimodal interaction*, Communication of the ACM, Vol. 42, No. 11, November, 1999, pp. 74-81.

Oviatt, S.L. (2003). *Multimodal interfaces*. In the Human-Computer Interaction Handbook : Fundamentals, Evolving Technologies and Emerging Applications, J. JACKO AND A. SEARS, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, 2003, chap.14, 286-304.

Tricot, A. & Chanquoy, L. (1996). La charge mentale, "vertu dormitive" ou concept opérationnel ? Introduction. *Psychologie Française*, 41 (4), 313-318.

Tricot, A. & Tricot, M. (2000). Un cadre formel pour interpréter les liens entre utilisabilité et utilité des systèmes d'information (et généralisation à l'évaluation d'objets finalisés). *Colloque Ergo-IHM 2000*, Biarritz, 3-6 octobre, communication orale.

## **8. Annexes**

Annexe 1 : le questionnaire d'évaluation subjective de la charge mentale

Annexe 2 : le questionnaire de satisfaction

Annexe 3 : grille d'entretien semi-directif sur la multimodalité

Annexe 4 : entretien "libre"

Annexe 5 : consignes

Annexe 6 : les scenarii

Annexe 7 : traitements statistiques sur le temps

Annexe 8 : traitements statistiques sur le nombre d'interactions

Annexe 9 : traitements statistiques sur l'indice de charge mentale

Annexe 10 : traitements statistiques sur l'indice de satisfaction





## Annexe 2 : le questionnaire de satisfaction

Pour chacune des phrases ci-dessous, vous allez vous situer sur une échelle en 5 points par rapport à chaque affirmation :

*Ainsi, vous vous positionnez de 1 si vous êtes totalement en désaccord avec l'affirmation, à 5 si vous êtes complètement d'accord avec l'affirmation*

Complètement en désaccord	Plutôt en désaccord	Mitigé	Plutôt d'accord	Complètement d'accord
1	2	3	4	5

1- J'ai été facilement compris(e) par le système	1	2	3	4	5
2- J'ai eu des difficultés pour formuler mes demandes	1	2	3	4	5
3- Le déplacement dans le service était simple	1	2	3	4	5
4- Je pouvais difficilement corriger une erreur dans le dialogue	1	2	3	4	5
5- Le service me guidait bien pour obtenir les renseignements qu'il attendait	1	2	3	4	5
6- J'ai été dérouté(e) par les réactions du système	1	2	3	4	5
7- J'ai obtenu des informations qui répondaient à mes attentes	1	2	3	4	5
8- J'ai trouvé difficilement les informations que je recherchais	1	2	3	4	5
9- Je ne savais pas à quel moment intervenir	1	2	3	4	5
10- J'ai trouvé gênante la quantité d'information délivrée	1	2	3	4	5

11- Je savais facilement comment intervenir	1	2	3	4	5
12- J'ai été dans l'ensemble satisfait par le service	1	2	3	4	5

13- J'ai trouvé l'interface conviviale et agréable	1	2	3	4	5
14- Je pense, dans l'avenir, utiliser ce service	1	2	3	4	5

### **Annexe 3 : grille d'entretien semi-directif sur la multimodalité**

#### Général :

1 - D'une manière générale, êtes-vous satisfait de la possibilité de pouvoir utiliser plusieurs modalités dans ce service (vocale, graphique, clic) ?

#### En entrée :

2 - Aimerez-vous pouvoir utiliser d'autres fonctions en vocal et en visuel (navigation, zoom sur la carte...)?

3 - Que pensez-vous de la façon dont on peut interagir vocalement avec l'interface ?

- un bouton push-to-talk (où vous devez maintenir un bouton appuyé pour parler, comme sur un talkie-walkie) vous semblerait-il plus approprié ?

- que penseriez-vous de la possibilité de parler de façon plus naturelle : sans avoir à appuyer sur un bouton pour prendre la parole ?

#### En sortie :

4 - Aimerez-vous avoir la possibilité de couper le son à tout moment ?

- la redondance synthèse vocale/informations graphiques a-t-elle constitué plutôt une aide ou plutôt une gêne pour vous ?

- plus précisément, dans quel(s) cas une aide et dans quel(s) cas une gêne ?

5 - Appréciez-vous le fait que les réponses soient données à la fois par la synthèse vocale et à l'écrit ?

6 - Comment avez-vous trouvé la synthèse vocale ? Avez-vous parfois eut envie de la couper pour pouvoir les lire ou préférez-vous uniquement les écouter ?

7 - Trouvez-vous ce service complet ? Aimerez-vous d'autres fonctionnalités, d'autres possibilités, d'autres contenus ?

8 - Avez-vous d'autres remarques ?

### ***Annexe 4 : entretien "libre"***

1 - Alors, comment ça s'est passé ? Comment vous avez vécu ces différentes recherches de restaurant avec Plan Resto ?

2 - Est-ce que vous avez pu faire tout ce que vous vouliez ?

3 - Quelles difficultés avez-vous rencontrées ? Vous êtes-vous senti à l'aise ?

4 - Y a-t-il des choses qui vous ont paru inutiles, gênantes ?

## **Annexe 5 : consignes**

### Versions multimodales :

L'objectif de ce test est d'évaluer un prototype du service Plan Resto d'un point de vue ergonomique.

Ce service interactif permet de rechercher un restaurant dans Paris intra-muros selon 3 critères : le lieu (arrondissement, station de métro, monument...), le prix, la spécialité.

*Néanmoins, vous ne pourrez pas effectuer de recherche à partir du nom d'un restaurant.*

Vous pourrez dialoguer avec Plan Resto en langage naturel (comme si vous vous adressiez à un opérateur humain) de 3 façons :

- soit en parlant dans le micro, sans oublier de cliquer sur le bouton "Cliquez ici pour parler"
- soit en tapant votre question dans le champ de saisie correspondant
- ou encore en cliquant sur les liens présents dans l'historique de votre dialogue avec Plan Resto

### **Nous ne testerons pas vos capacités mais celles du service !**

Le test se déroulera de la manière suivante : après un premier essai où vous serez libre, je vous présenterai des scénarios de recherche (comprenant un ou plusieurs critères) et vous devrez réaliser ces recherches en vous projetant le plus possible dans la situation décrite. Néanmoins, vous n'irez pas jusqu'à être mis en relation avec le restaurant.

Je vous demanderai également de remplir un petit questionnaire à la fin de chaque scénario.

Le service étant encore à l'état de prototype, il se peut qu'il y ait des bugs. Dans ce cas, j'interviendrai pour rétablir le service.

Avez-vous des questions avant de commencer ?

Avant le 1<sup>er</sup> scénario, dire : "dites-moi quand vous êtes prêt et je vous dirai quand vous pourrez commencer [le temps de déclencher l'enregistrement vidéo]" et "une fois que vous avez obtenu les informations dont vous avez besoin, dites : 'j'ai terminé' "

Version graphique :

L'objectif de ce test est d'évaluer un prototype du service Plan Resto d'un point de vue ergonomique.

Ce service interactif permet de rechercher un restaurant dans Paris intra-muros selon 3 critères : le lieu (arrondissement, station de métro, monument...), le prix, la spécialité.

*Néanmoins, vous ne pourrez pas effectuer de recherche à partir du nom d'un restaurant.*

Vous pourrez dialoguer avec Plan Resto en langage naturel (comme si vous vous adressiez à un opérateur humain) de 2 façons :

- soit en tapant votre question dans le champ de saisie correspondant
- soit en cliquant sur les liens présents dans l'historique de votre dialogue avec Plan Resto

**Nous ne testerons pas vos capacités mais celles du service !**

Le test se déroulera de la manière suivante : après un premier essai où vous serez libre, je vous présenterai des scénarios de recherche (comprenant un ou plusieurs critères) et vous devrez réaliser ces recherches en vous projetant le plus possible dans la situation décrite. Néanmoins, vous n'irez pas jusqu'à être mis en relation avec le restaurant.

Je vous demanderai également de remplir un petit questionnaire à la fin de chaque scénario.

Le service étant encore à l'état de prototype, il se peut qu'il y ait des bugs. Dans ce cas, j'interviendrai pour rétablir le service.

Avez-vous des questions avant de commencer ?

Avant le 1<sup>er</sup> scénario, dire : "dites-moi quand vous êtes prêt et je vous dirai quand vous pourrez commencer [le temps de déclencher l'enregistrement vidéo]" et "une fois que vous avez obtenu les informations dont vous avez besoin, dites : 'j'ai terminé' "

Version vocale :

L'objectif de ce test est d'évaluer un prototype du service Plan Resto d'un point de vue ergonomique.

Ce service interactif permet de rechercher un restaurant dans Paris intra-muros selon 3 critères : le lieu (arrondissement, station de métro, monument...), le prix, la spécialité.

*Néanmoins, vous ne pourrez pas effectuer de recherche à partir du nom d'un restaurant.*

Vous pourrez dialoguer avec Plan Resto en langage naturel (comme si vous vous adressiez à un opérateur humain)

**Nous ne testerons pas vos capacités mais celles du service !**

Le test se déroulera de la manière suivante : après un premier essai où vous serez libre, je vous présenterai des scénarios de recherche (comprenant un ou plusieurs critères) et vous devrez réaliser ces recherches en vous projetant le plus possible dans la situation décrite. Néanmoins, vous n'irez pas jusqu'à être mis en relation avec le restaurant.

Je vous demanderai également de remplir un petit questionnaire à la fin de chaque scénario.

Le service étant encore à l'état de prototype, il se peut qu'il y ait des bugs. Dans ce cas, j'interviendrai pour rétablir le service.

Avez-vous des questions avant de commencer ?

Avant le 1<sup>er</sup> scénario, dire : "dites-moi quand vous êtes prêt et je vous dirai quand vous pourrez commencer [le temps de déclencher l'enregistrement vidéo]" et "une fois que vous avez obtenu les informations dont vous avez besoin, dites : 'j'ai terminé' "

## **Annexe 6 : les scenarii**

### **Scénario n° 1 : essai libre**

Vous devez trouver un restaurant dans Paris intra-muros. Vous êtes libre de choisir les critères de recherche que vous voulez (lieu, prix, spécialité).

Néanmoins, une fois que vous vous êtes fixé vos critères, nous vous demandons de bien vouloir les conserver jusqu'au bout de votre recherche.

Effectuez votre recherche en essayant d'obtenir le maximum d'informations, comme vous le feriez en situation réelle.

=> voir comment l'utilisateur prend en main le service spontanément

### **Scénario n° 2 : tâche simple**

Un ami vous propose un déjeuner sur Paris. Il vous a parlé du service Plan Resto pour vous aider à trouver un restaurant gastronomique près du Trocadéro. Il ne vous donne pas plus d'information.

Vous interrogez le service afin de trouver un restaurant avec les critères que vous vous êtes fixés. A vous de préciser **ensuite** d'autres critères si besoin.

Effectuez votre recherche en essayant d'obtenir le maximum d'informations, comme vous le feriez en situation réelle.

=> tâche simple sur le choix de solutions trouvées par le système : 1 seule solution explicite  
2 critères spécifiés : une spécialité et un lieu précis

### **Scénario n° 3 : tâche moyennement difficile**

Vous êtes dans le 15ème arrondissement. Vous voulez absolument manger dans un restaurant végétarien.

Vous seriez content de trouver dans cet arrondissement, sinon à voir éventuellement quels autres arrondissements proposent cette spécialité, vous êtes prêt à marcher.

Vous interrogez le service afin de trouver un restaurant avec les critères que vous vous êtes fixés. A vous de préciser ensuite d'autres critères si besoin.

Effectuez votre recherche en essayant d'obtenir le maximum d'informations, comme vous le feriez en situation réelle.

=> tâche complexe : il n'y a pas de restaurant végétarien dans le 15<sup>ème</sup> arrondissement inclus dans la base de données de Plan Resto, une relaxation est faite sur des solutions alternatives.

2 critères spécifiés : un arrondissement et une spécialité

#### **Scénario n° 4 : tâche moyennement difficile**

Après une journée où vous avez beaucoup marché, vous souhaitez dîner près du métro Filles du calvaire, sans idée précise de spécialité, ni de prix au départ.

Vous interrogez le service afin de trouver un restaurant avec les critères que vous vous êtes fixés. A vous de préciser ensuite d'autres critères si besoin.

Effectuez votre recherche en essayant d'obtenir le maximum d'informations, comme vous le feriez en situation réelle.

=> tâche "piège", la reconnaissance vocale ne reconnaît pas "Filles du calvaire", il n'y a qu'à l'écrit que cette requête peut être correctement prise en compte. De plus, le contexte "une journée où vous avez beaucoup marché" sous entend que l'utilisateur devrait inférer le fait qu'il doit trouver une restaurant proche du métro "Filles du calvaire"

1 critère spécifié : une station de métro

N.B. : Pour la version vocale, ce scénario a été modifié :

Après une journée où vous avez beaucoup marché, vous souhaitez dîner près du métro Alésia, sans idée précise de spécialité, ni de prix au départ.

Vous interrogez le service afin de trouver un restaurant avec les critères que vous vous êtes fixés. A vous de préciser ensuite d'autres critères si besoin.

Effectuez votre recherche en essayant d'obtenir le maximum d'informations, comme vous le feriez en situation réelle.

## Annexe 7 : traitements statistiques sur le temps

### SUR L'ENSEMBLE DES 3 SCENARI

ANOVA effet du facteur Version (vocale, graphique, multimodale sans avatar et multimodale avec avatar) sur l'ensemble des temps (bruts) pour les 3 premiers scénarii :

Tests Multivariés de Significativité (Feuille de données Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace)						
Effet	Test	Valeur	F	Effet dl	Erreur dl	p
ord. origine	Wilk	0,084734	64,80972	3	18,00000	0,000000
VERSION	Wilk	0,452127	1,88354	9	43,95788	0,079883

=> NS

### SCENARIO N° 1

Distribution des temps bruts pour le scénario n°1 :

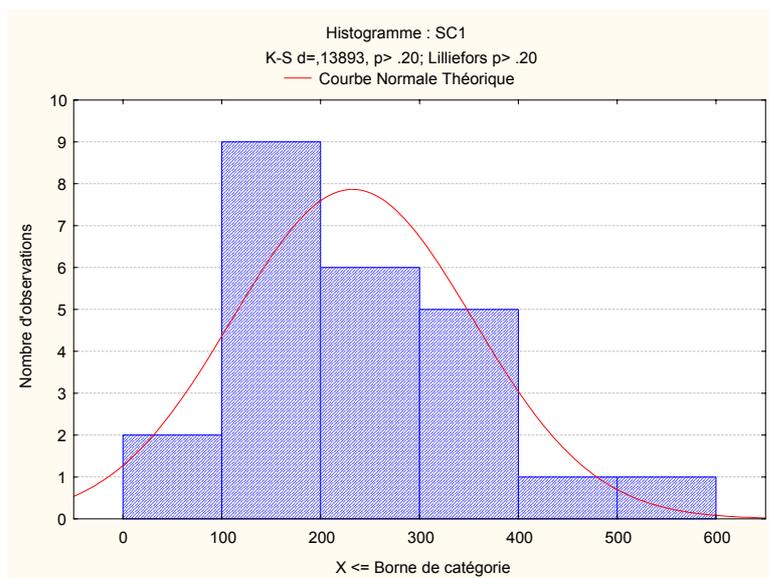


Tableau résumé stat des temps bruts pour tous le scénario n°1 :

Variable	Statistiques Descriptives (Feuille de données1)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SC1	24	232,2500	218,0000	52,00000	521,0000	469,0000	121,7068

ANOVA effet du facteur Version (vocale, graphique, multimodale sans avatar et multimodale avec avatar) sur les temps bruts au scénario n°1 :

Effet	Tests Univariés de Significativité pour SC1 (Feuille Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace)				
	SC	Degr. de Liberté	MC	F	p
ord. origine	1294562	1	1294562	86,22966	0,000000
VERSION	40430	3	13477	0,89766	0,459654
Erreur	300259	20	15013		

=> NS

## SCENARIO N° 2

Distribution des temps bruts pour le scénario n°2 :

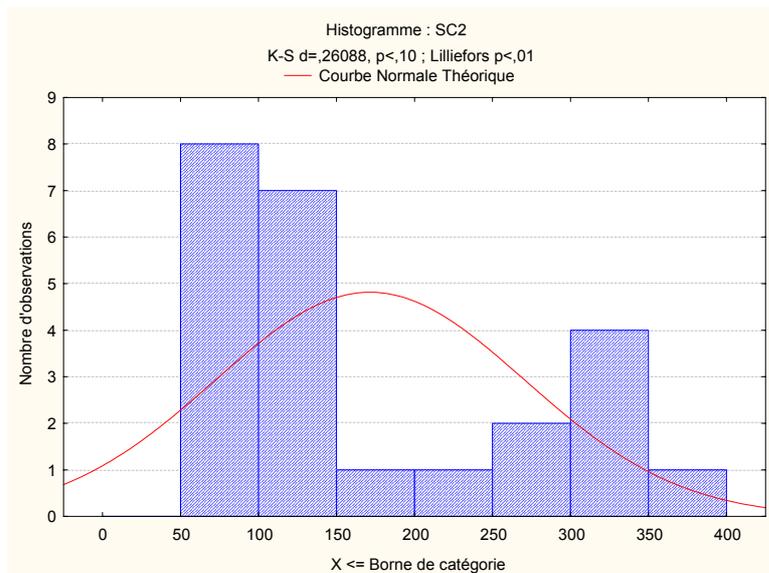


Tableau résumé stat des temps bruts pour tous le scénario n°2 :

Statistiques Descriptives (Feuille de données1)							
Variable	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SC2	24	171,5417	121,0000	55,00000	359,0000	304,0000	99,40824

ANOVA effet du facteur Version (vocale, graphique, multimodale sans avatar et multimodale avec avatar) sur les temps bruts au scénario n°2 :

Tests Univariés de Significativité pour SC2 (Feuille d Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace					
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	706237,0	1	706237,0	73,96889	0,000000
VERSION	36330,8	3	12110,3	1,26839	0,312118
Erreur	190955,2	20	9547,8		

=> NS

### SCENARIO N° 3

Distribution des temps bruts pour le scénario n°3 :

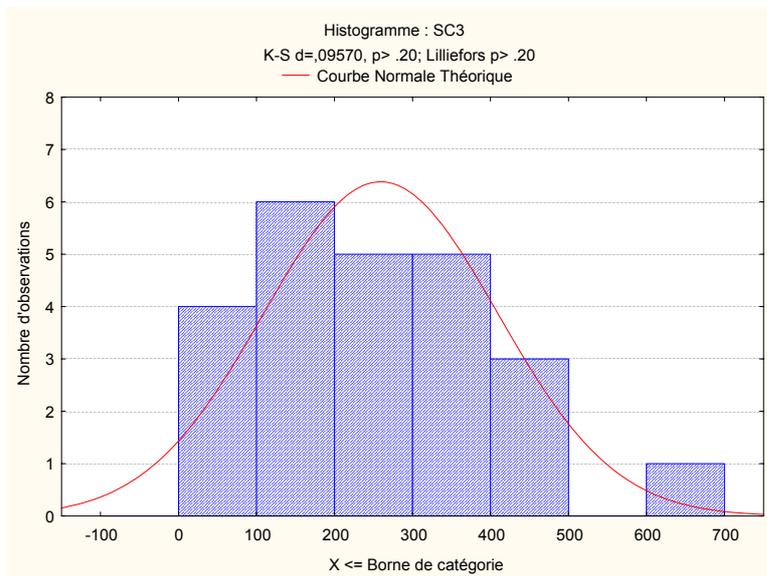


Tableau résumé stat des temps bruts pour tous le scénario n°3 :

Variable	Statistiques Descriptives (Feuille de données1)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SC3	24	259,3333	252,0000	46,00000	650,0000	604,0000	149,9985

ANOVA effet du facteur Version (vocale, graphique, multimodale sans avatar et multimodale avec avatar) sur les temps bruts au scénario n°3 :

Effet	Tests Univariés de Significativité pour SC3 (Feuille Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace				
	SC	Degr. de Liberté	MC	F	p
ord. origine	1614091	1	1614091	101,8858	0,000000
VERSION	200646	3	66882	4,2218	0,018213
Erreur	316843	20	15842		

=> il y a un effet du facteur Version sur le temps pour effectuer le scénario n°3

Comparaisons partielles :

MM sans avatar VS MM avec avatar

Source	Test Univarié de Significativité pour les Comp. Plann Variable dépendante : SC3				
	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	11285,3	1	11285,33	0,712361	0,408643
Erreur	316843,0	20	15842,15		

=> NS

Vocale VS Graphique

Test Univarié de Significativité pour les Comp. Plann					
Variable dépendante : SC3					
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	7008,3	1	7008,33	0,442385	0,513571
Erreur	316843,0	20	15842,15		

=> NS

Vocale VS MM sans avatar :

Test Univarié de Significativité pour les Comp. Plann					
Variable dépendante : SC3					
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	84504,1	1	84504,08	5,334130	0,031711
Erreur	316843,0	20	15842,15		

=> les sujets de la version vocale mettent significativement plus de temps pour effectuer le scénario n°3 que les sujets de la version multimodale sans avatar.

Vocale VS MM avec avatar :

Test Univarié de Significativité pour les Comp. Plann					
Variable dépendante : SC3					
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	157552,1	1	157552,1	9,945120	0,004998
Erreur	316843,0	20	15842,2		

=> les sujets de la version vocale mettent significativement plus de temps pour effectuer le scénario n°3 que les sujets de la version multimodale avec avatar.

Graphique VS MM avec :

Test Univarié de Significativité pour les Comp. Plann					
Variable dépendante : SC3					
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	98102,1	1	98102,08	6,192473	0,021762
Erreur	316843,0	20	15842,15		

=> les sujets de la version graphique mettent significativement plus de temps pour effectuer le scénario n°3 que les sujets de la version multimodale avec avatar.

Graphique VS MM sans :

Test Univarié de Significativité pour les Comp. Plann					
Variable dépendante : SC3					
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	42840,7	1	42840,75	2,704226	0,115710
Erreur	316843,0	20	15842,15		

=> NS

Vocal + Graphique VS MM sans + MM avec :

Test Univarié de Significativité pour les Comp. Plann					
Variable dépendante : SC3					
Source	Somme Carrés	Degr. de Liberté	Moyen. Carré	F	p
Effet	182352,7	1	182352,7	11,51060	0,002889
Erreur	316843,0	20	15842,2		

=> les sujets des versions monomodales mettent moins de temps pour effectuer le scénario n°3 que les sujets des versions multimodales

## Annexe 8 : traitements statistiques sur le nombre d'interactions

### SUR L'ENSEMBLE DES 3 SCENARII

Distribution nombre d'interactions bruts sur les 3 premiers scenarii

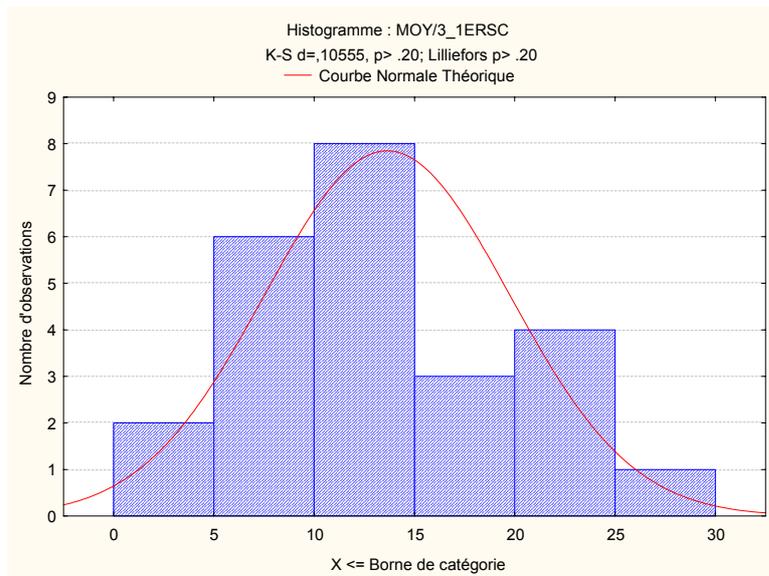


Tableau résumé stat nombre d'interactions bruts sur les 3 premiers scenarii :

Variable	Statistiques Descriptives (donees_tour-parole)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
MOY/3_1ERSC	24	13,63889	12,50000	3,666667	27,33333	23,66667	6,102123

ANOVA effet du facteur version sur le nombre d'interactions bruts sur les 3 premiers scenarii :

Effet	Tests Multivariés de Significativité (donees_tour-parole) Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace					
	Test	Valeur	F	Effet dl	Erreur dl	p
ord. origine	Wilk	0,116786	45,37615	3	18,00000	0,000000
VERSION	Wilk	0,653949	0,93127	9	43,95788	0,508027

=> NS

## SCENARIO 1

Distribution du nombre d'interactions pour le scénario n°1 :

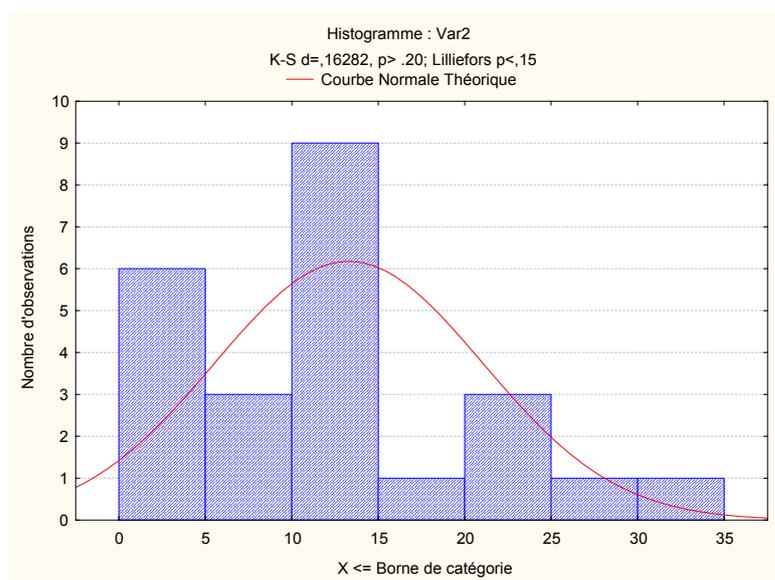


Tableau résumé stat pour le nombre d'interactions pour le scénario n°1 :

Variable	Statistiques Descriptives (Feuille de données1)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
Var2	24	13,29167	12,00000	3,000000	32,00000	29,00000	7,754265

Anova effet du facteur Versions sur le nombre d'interactions pour le scénario n°1

Tests Univariés de Significativité pour Var2 (Feuille de Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace					
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	4240,042	1	4240,042	74,72536	0,000000
"Var1"	248,125	3	82,708	1,45763	0,256114
Erreur	1134,833	20	56,742		

=> NS

## SCENARIO 2

Distribution du nombre d'interactions pour le scénario n°2 :

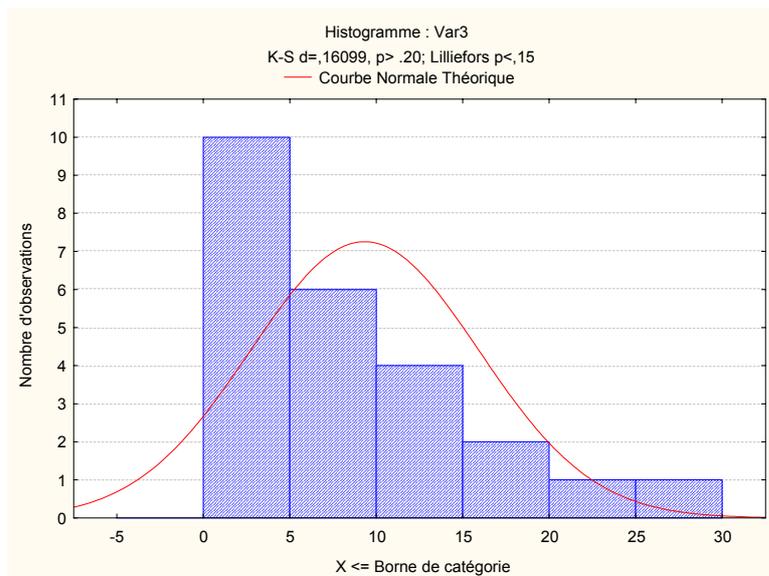


Tableau résumé stat pour le nombre d'interactions pour le scénario n°2 :

Variable	Statistiques Descriptives (Feuille de données1)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
Var3	24	9,333333	8,000000	1,000000	26,00000	25,00000	6,598199

Anova effet du facteur Versions sur le nombre d'interactions pour le scénario n°2

Tests Univariés de Significativité pour Var3 (Feuille de Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace					
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	2090,667	1	2090,667	45,41636	0,000001
"Var1"	80,667	3	26,889	0,58412	0,632355
Erreur	920,667	20	46,033		

=> NS

### SCENARIO 3

Distribution du nombre de tours de parole sur le scénario n°3 :

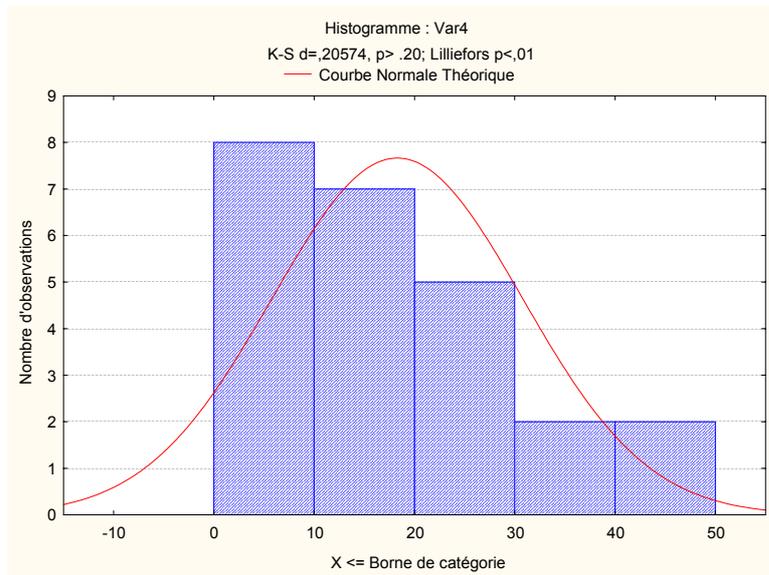


Tableau résumé stat pour le nombre d'interactions pour le scénario n°3 :

Variable	Statistiques Descriptives (Feuille de données1)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
Var4	24	18,29167	13,00000	4,000000	49,00000	45,00000	12,49166

Anova effet du facteur Versions sur le nombre d'interactions pour le scénario n°3

Effet	Tests Univariés de Significativité pour Var4 (Feuille de Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace				
	SC	Degr. de Liberté	MC	F	p
ord. origine	8030,042	1	8030,042	54,35191	0,000000
"Var1"	634,125	3	211,375	1,43071	0,263408
Erreur	2954,833	20	147,742		

=> NS

## Annexe 9 : traitements statistiques sur l'indice de charge mentale

### SUR L'ENSEMBLE DES 3 SCENARIII

Anova effet du facteur Version sur l'ensemble des indices de charge mentale pour les 3 premiers scénariii :

Tests Multivariés de Significativité (donnees_charge-m Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace						
Effet	Test	Valeur	F	Effet dl	Erreur dl	p
ord. origine	Wilk	0,058023	97,40679	3	18,00000	0,000000
VERSIONS	Wilk	0,657289	0,91911	9	43,95788	0,517860

=> NS

### SCENARIO 1

Distribution de l'indice de charge mentale sur le scénario n°1 :

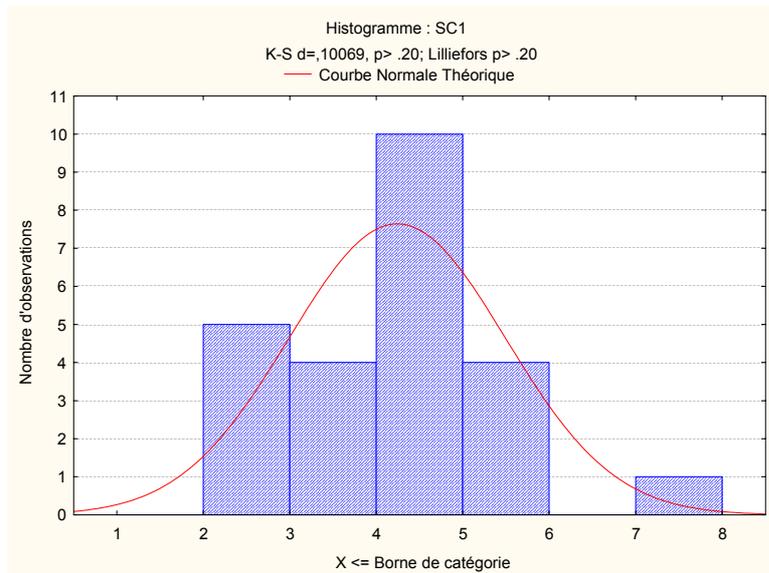


Tableau résumé stat pour l'indice de charge mentale sur le scénario n°1 :

Statistiques Descriptives (Feuille de données18)							
Variable	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SC1	24	4,243056	4,333333	2,333333	7,666667	5,333333	1,252996

Anova effet du facteur Versions sur l'indice de charge mentale pour le scénario n°1 :

Tests Univariés de Significativité pour SC1 (Feuille d Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace)					
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	432,0845	1	432,0845	296,8047	0,000000
VERSIONS	6,9942	3	2,3314	1,6015	0,220523
Erreur	29,1157	20	1,4558		

=> NS

## SCENARIO 2

Distribution de l'indice de charge mentale sur le scénario n°2 :

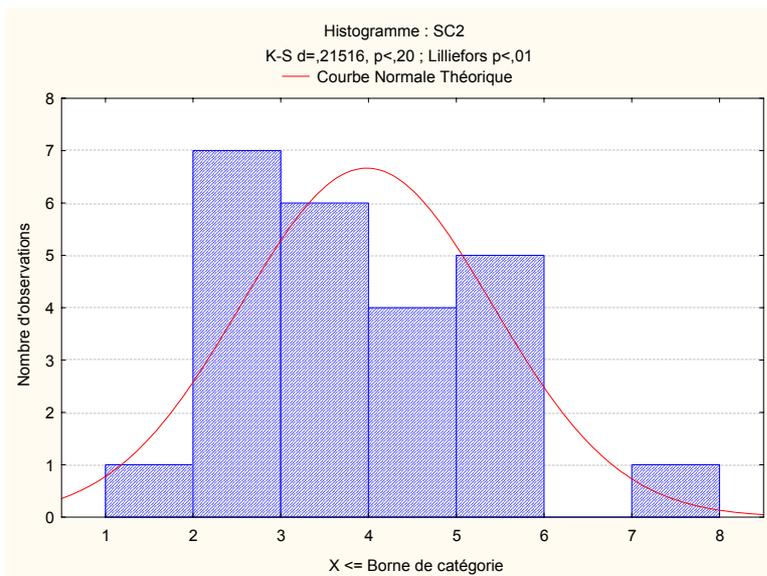


Tableau résumé stat pour l'indice de charge mentale sur le scénario n°2 :

Statistiques Descriptives (Feuille de données18)							
Variable	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SC2	24	3,979167	3,333333	1,666667	7,666667	6,000000	1,436509

Anova effet du facteur Versions sur l'indice de charge mentale pour le scénario n°2 :

Tests Univariés de Significativité pour SC2 (Feuille d Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace					
Effet	SC	Degr. de Liberté	MC	F	p
ord. origine	380,0104	1	380,0104	167,4977	0,000000
VERSIONS	2,0868	3	0,6956	0,3066	0,820304
Erreur	45,3750	20	2,2688		

=> NS

### SCENARIO 3

Distribution de l'indice de charge mentale sur le scénario n°3 :

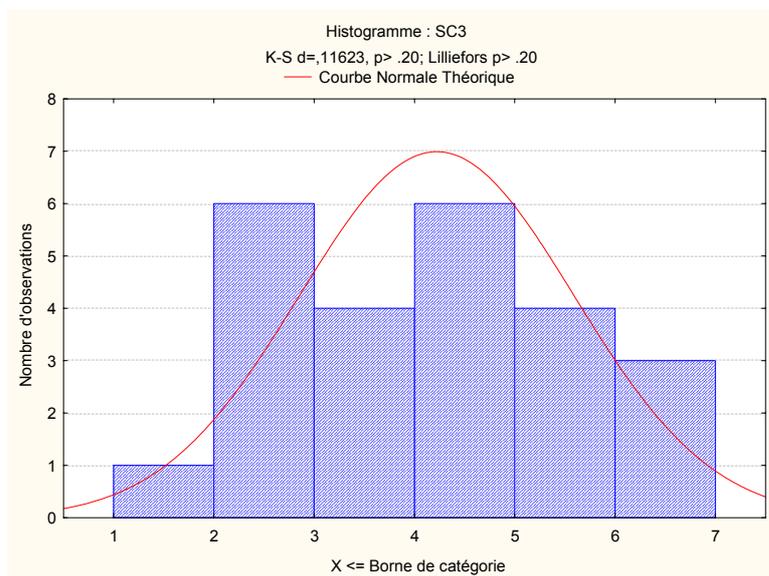


Tableau résumé stat pour l'indice de charge mentale sur le scénario n°3 :

Statistiques Descriptives (Feuille de données18)							
Variable	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SC3	24	4,222222	4,333333	1,666667	6,500000	4,833333	1,369674

Anova effet du facteur Versions sur l'indice de charge mentale pour le scénario n°3 :

Effet	Tests Univariés de Significativité pour SC3 (Feuille d Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace				
	SC	Degr. de Liberté	MC	F	p
ord. origine	427,8519	1	427,8519	217,6030	0,000000
VERSIONS	3,8241	3	1,2747	0,6483	0,593174
Erreur	39,3241	20	1,9662		

=> NS

## Annexe 10 : traitements statistiques sur l'indice de satisfaction

Distribution de l'indice de satisfaction sur l'ensemble des sujets :

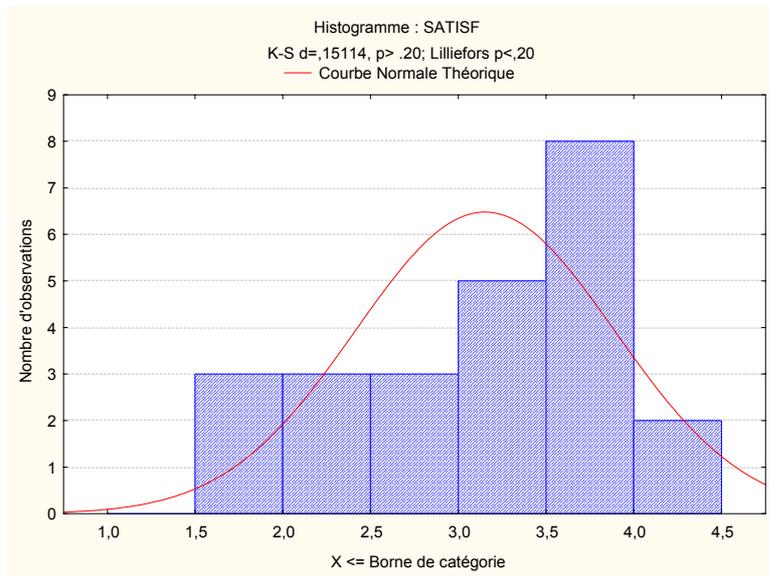


Tableau résumé stat de l'indice de satisfaction pour l'ensemble des sujets :

Variable	Statistiques Descriptives (Feuille de données32)						
	N Actifs	Moyenne	Médiane	Minimum	Maximum	Etendue	Ecart-type
SATISF	24	3,151786	3,357143	1,714286	4,214286	2,500000	0,738807

Anova effet du facteur Version sur l'indice de satisfaction :

Effet	Tests Univariés de Significativité pour SATISF (Feuil Paramétrisation sigma-restreinte Décomposition de l'hypothèse efficace)				
	SC	Degr. de Liberté	MC	F	p
ord. origine	238,4101	1	238,4101	409,5089	0,000000
VERSION	0,9105	3	0,3035	0,5213	0,672509
Erreur	11,6437	20	0,5822		

=> NS