



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 11592

To link to this article : DOI: 10.1145/1148550.1148600
URL : <http://dx.doi.org/10.1145/1148550.1148600>

To cite this version : Karsenty, Laurent and Sire, Stéphane and Cause, Mickael and Deherly, Nicolas *Quel impact de l'entrée vocale sur la conception graphique d'un service mobile ?* (2005) In: IHM 2005, 27 September 2005 - 30 September 2005 (Toulouse, France).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Quel impact de l'entrée vocale sur la conception graphique d'un service mobile ?

Laurent Karsenty, Stéphane Sire, Mickael Causse, Nicolas Deherly

IntuiLab
BP 27201
Prologue 1 – La Pyrénéenne
31672 Labège, Cedex, France
{karsenty, sire, causse, deherly}@intuilab.com

RESUME

Le développement d'interfaces multimodales bénéficiant de l'entrée vocale se traduira dans un certain nombre de cas par une évolution d'une conception graphique existante. Entre adaptation mineure et reconception totale, se pose la question de savoir quel impact une telle évolution sur l'existant peut avoir. Pour apporter des éléments de réponse à cette question, une expérience a été menée dans le cadre du projet européen Use-Me.Gov pour lequel on devait reconcevoir un service existant de diffusion d'information sur des événements municipaux en y ajoutant une entrée vocale. Cet article rapporte les principaux résultats de cette expérience ainsi que les enseignements que nous en avons tirés.

MOTS CLES : Multimodalité, Voix, Langue Naturelle, Mobilité, Conception.

ABSTRACT

The development of multimodal interfaces exploiting voice input will appear in a number of cases as an evolution of an existing graphic design. Between minor adaptation and full re-design, the question arises as to which impact such an evolution could have on existing user interfaces. To bring initial answers an experiment was conducted within the European project Use-Me.Gov in which our goal was to redesign an existing service broadcasting news on municipal events by adding natural language voice input to it. This communication reports the main results of this experiment as well as the lessons that we drew from it.

CATEGORIES AND SUBJECT DESCRIPTORS: H.5.2 [Information Interfaces and Presentation]: User Interfaces---Screen Design, Voice I/O, Natural Language.

GENERAL TERMS: Design

KEYWORDS: Multimodality, Voice, Natural Language, Mobility, Design.

INTRODUCTION

Les interfaces multimodales sont étudiées dans les laboratoires de recherche depuis plus d'une quinzaine d'années [6]. Aujourd'hui, grâce à la maturation d'un certain nombre de technologies d'interaction et à la demande de certains domaines d'application, en particulier le domaine de la mobilité, ces produits de la recherche sont en passe de devenir des produits réels. Dans un certain nombre de cas, leur développement risque fort de résulter d'une évolution des applications existantes plutôt que d'une définition totalement nouvelle. Se pose dès lors la question de savoir quel impact peut avoir l'ajout de nouvelles modalités sur l'application existante, et en particulier sur la conception de son interface utilisateur.

Cette question a été traitée dans le cadre d'un projet européen, Use-Me.Gov, que nous décrivons plus en détail dans la section suivante. Nous y avons mené une expérience visant à adapter une interface graphique déjà conçue pour ajouter une entrée vocale. L'objectif de cette expérimentation était de dégager des principes d'adaptation de l'interface existante et d'aider les décideurs et concepteurs à anticiper les effets de la « multimodalisation » de services existants sur le processus de conception. Comment modifier le service pour favoriser l'entrée vocale ? Doit-on s'attendre à quelques adaptations relativement légères ou doit-on anticiper une reconception globale du service ? Telles sont quelques questions qui ont guidé notre étude. Cet article en rapporte les principaux résultats et propose de tirer quelques leçons de cette expérience, en attendant de pouvoir la répéter et ainsi d'en vérifier le caractère plus ou moins général.

CONTEXTE DU PROJET USE-ME.GOV

La mise en place des infrastructures de télécommunication mobile et leur interconnexion avec le réseau Internet offre la possibilité de créer des services qui s'adressent à tous les citoyens possesseurs d'un téléphone mobile. Ce nouveau type de déploiement s'accompagne de contraintes fortes sur la conception des interfaces Homme-Machine : essentiellement conception

pour petits écrans et conception pour des utilisateurs non familiers avec l'informatique.

Le projet USE-ME.GOV est un projet IST-2002 de la commission européenne destiné à faciliter la diffusion des services mobiles aux citoyens. Un premier volet de ce projet, qui n'est pas traité dans cet article, concerne la création d'une plateforme logicielle d'interopérabilité entre les services informatiques des autorités locales (villes, chambres de commerce, hôpitaux, etc.) et les plateformes des opérateurs téléphoniques. Le second volet du projet concerne la recherche et l'étude de recommandations pour l'adaptation des interfaces homme-machine aux différentes caractéristiques des téléphones mobiles des utilisateurs.

Dans ce contexte nous avons mené une expérience de re-conception pour passer d'un environnement monomodal tactile à un environnement multimodal tactile et vocal. Notre objectif était également d'améliorer au passage l'intuitivité de l'interface pour des utilisateurs novices grâce à l'apport de la multimodalité. Le service et son interface, tel qu'ils ont été définis initialement, sont présentés dans la section suivante.

PRESENTATION DU SERVICE

Le service retenu pour cette étude est un service de diffusion d'information sur des événements municipaux. Il s'adresse aux citoyens mobiles munis d'un téléphone. La ville de Paris a été choisie à titre d'exemple. La maquette du service qui nous a servi de point départ et qui illustre cette section est la version obtenue après intervention des graphistes du partenaire espagnol Indra sur une maquette initialement créée par le Fraunhofer Institute de Bonn.

Fonctionnalités

Le service a pour but de filtrer des événements concernant la vie locale de la cité. Les événements sont ensuite envoyés sous forme de SMS ou de MMS. Les événements sont groupés par canaux :

- Culture: événements culturels comme les annonces de spectacle ;
- Expos : événements commerciaux comme les salons et les foires ;
- Circulation : événements liés aux conditions de trafic ;
- Sport : événements sportifs.

Le service offre deux modes d'accès:

- recherche immédiate
- abonnement et filtrage

Dans le mode de recherche immédiate, illustré par la Figure 1, l'utilisateur sélectionne un canal, une ou plusieurs périodes de temps et un ou plusieurs arrondissements pour effectuer une interrogation. Par exemple, il peut demander : « des informations sur la

circulation dans le 19ème arrondissement pour demain ». Après affichage d'un écran de confirmation, le système indique le nombre de messages (SMS ou MMS) qui sont aussitôt envoyés.

Dans le mode d'abonnement et de filtrage, l'utilisateur a la possibilité de définir un abonnement pour chacun des canaux. L'abonnement est associé avec des préférences qui servent à filtrer les messages reçus par rapport à une semaine type et le moment de réception. Par exemple l'utilisateur peut s'abonner aux événements sportifs se déroulant tous les samedi en étant prévenu une fois par semaine.



Figure 1: Requête multi-critère pour la recherche immédiate à gauche avec sélection des bornes de la période à droite.¹

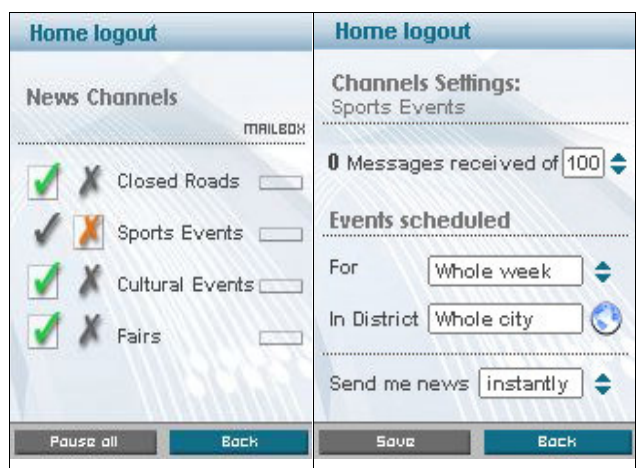


Figure 2: Requête simple pour l'abonnement ou le désabonnement à gauche et requête multi-critères pour fixer les préférences d'un abonnement à droite.

Besoins utilisateurs

La maquette tactile initiale créée par Fraunhofer Institute et les partenaires du projet, en particulier la Municipalité de Bologne, a été conçue en fonction d'une analyse des besoins qui précisait que l'usage attendu était peu

¹ L'Anglais est utilisé sur cette maquette d'interface car elle a été réalisée pour alimenter les spécifications du service pilote qui, lui, a été réalisé en Italien et testé à Bologne.

fréquent, les utilisateurs non familiers avec l'informatique et l'usage mobile.

Le besoin de répondre à un usage peu fréquent implique de renforcer la facilité d'apprentissage et le guidage. Ce besoin se traduit par un choix délibéré de fonctionnalités simples. Ainsi, pour la sélection de la période de notification lors de l'abonnement à un canal, les trois critères possibles n'ont pas de paramètre supplémentaire. Par exemple, la notification quotidienne a lieu tous les jours à une heure définie par le service qui n'est pas modifiable. Ce besoin se traduit aussi par des pages écran sans surcharge, à l'instar de la page d'accueil montrée sur la Figure 3 qui ne contient que le choix entre les deux modes d'accès : recherche immédiate ou abonnement, avec une variété chromatique simple et une sobriété des graphismes.

Le second besoin à prendre en compte, utilisateurs non familiers, se traduit par un lexique simple pour les labels et le doublage des icônes par des textes descriptifs. L'interaction est en outre limitée au pointage et les différentes pages du service sont homogènes.

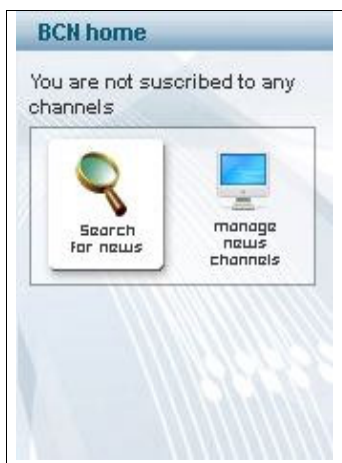


Figure 3: Page d'accueil du service, les icônes sont doublées d'un texte pour satisfaire aux besoins utilisateurs et la conception graphique est dépouillée.

Le troisième besoin – usage mobile - a été pris en compte sur la page d'accueil qui contient seulement deux composants d'interaction relativement larges pour faciliter la sélection au doigt. L'ajout de la modalité vocale prend en compte ce besoin. L'interaction par la voix est en effet un moyen d'entrée accessible si l'utilisateur n'est pas en situation pour pointer convenablement son écran tactile, par exemple au cours de la locomotion.

Caractéristiques du service

Le service est assimilable à un service d'édition d'une requête multi-critères. Dans le cas de la recherche immédiate d'événements, les critères sont le canal d'information, la période de temps indiquée par un jour de début et un jour de fin et la liste des arrondissements concernés. Dans le cas de la modification des préférences

d'abonnement à un canal, les critères sont la période de temps sur la semaine concernée avec un choix entre les jours de la semaine, la liste des arrondissements concernés, et le moment de la notification avec un choix entre une notification instantanée, quotidienne ou hebdomadaire. La notification instantanée signifie que l'utilisateur reçoit un message dès qu'un nouvel événement correspondant à ses critères est signalé par une des autorités locales.

En plus de ces deux requêtes complexes, le service comporte également deux requêtes simples pour s'abonner à un canal ou bien se désabonner d'un canal.

Navigation

Le diagramme de navigation dans le service est représenté sur la Figure 4. La maquette comporte 9 pages. La profondeur maximale est de 5 pages.

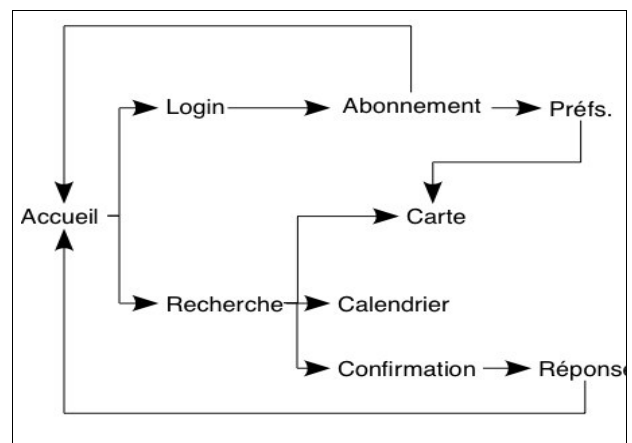


Figure 4: Diagramme de navigation entre pages

METHODE DE CONCEPTION

La conception est partie de la maquette existante et d'un document de spécification des besoins. Un groupe pluridisciplinaire de conception a été formé comprenant ergonomes, informaticiens spécialisés en interaction homme-machine et un designer graphique. Le groupe de conception a travaillé selon une *démarche itérative et participative* adaptée aux contraintes d'un projet, afin d'assurer la réalisation d'un prototype satisfaisant au mieux et au plus vite les besoins utilisateurs et les contraintes d'utilisabilité connus. Cette démarche a déjà été présentée à IHM'04 [7] et ne sera pas reprise en détail ici. Seulement certaines étapes spécifiques seront mentionnées.

Ainsi, la reconception est partie d'une évaluation de la maquette existante basée sur la méthode d'inspection cognitive [9] et la rédaction de scénarios d'utilisation. Différents scénarios ont été envisagés et, très vite, ceux basés sur l'utilisation de l'entrée vocale sont apparus comme les plus prometteurs en termes d'utilisabilité et de compatibilité avec les contraintes de la mobilité tout en étant réalistes en termes de faisabilité et de

déploiement dans un avenir relativement proche (d'ici 2 à 3 ans environ).

Sur la base des scénarios retenus, plusieurs problèmes de conception ont été identifiés et des séances de conception participative ont été mises en place pour les traiter. Pour la seule page d'accueil, près d'une dizaine de solutions ont été explorées. Certaines ont pu être éliminées sur la base d'une grille de critères. Ensuite, à partir de dessins d'écran, quatre d'entre elles ont été testées du point de vue de la compréhension auprès d'un petit groupe d'utilisateurs. L'une d'elles est alors apparue comme la plus facile à comprendre et a été retenue (voir figure 6).

Une première maquette de l'interface multimodale a été produite. Elle a ensuite été évaluée, toujours en exploitant une méthode d'inspection cognitive, par des collègues spécialistes en utilisabilité du Fraunhofer Institute de Bonn. Cette évaluation a produit plus d'une trentaine de remarques qui ont été intégralement prises en compte dans la réalisation du prototype du service multimodal. Dans une phase ultérieure, prévue pour le second trimestre 2005, un test utilisateur va être mis en place.

ANALYSE DE L'IMPACT DE L'ENTREE VOCALE

L'objectif de la conception de l'interface multimodale est de comprendre par la pratique quelles peuvent être les conséquences possibles de l'ajout d'une modalité vocale sur une interface déjà conçue indépendamment de la voix. La description de ces conséquences, répertoriées dans la suite à travers différents thèmes, est présentée après quelques remarques sur le prototype réalisé et le choix du type de langage vocal que nous avons fait.

Choix du type de langage vocal

Le prototype réalisé simule un smartphone de type P900 sur un iPaq tout respectant sa taille et sa résolution. Il propose à tout moment la possibilité d'interagir au stylet (ou au doigt) et/ou à la voix. Son développement, réalisé avec notre environnement logiciel IntuiKit [2], s'est appuyé sur le moteur de reconnaissance de parole de Nuance. S'est alors posée la question du type de langage que nous autorisons.

Une première possibilité consistait à employer des commandes vocales par mots-clés pour remplir les champs de chaque page et pour naviguer de page en page. Cette approche a l'avantage d'être rapide et plus ou moins automatique car elle ne change pas la structure de l'interface ni la présentation des informations. Cependant, à notre connaissance, rien ne vient appuyer l'hypothèse que, dans ces conditions, l'ajout de la modalité vocale apporterait un gain d'efficacité ou de facilité d'apprentissage. Il faut noter que l'utilisation de la voix sur dispositif mobile impose généralement de déclencher la reconnaissance de parole qui ne peut rester ouverte en permanence (Push-to-Talk). Avec un smartphone, cela revient à enclencher un bouton physique ou cliquer sur un bouton d'écran. Or, comme les mots clés

correspondent assez souvent à des boutons ou des items présentés à l'écran, l'avantage que pourrait avoir la voix disparaît assez souvent (sauf à envisager des contextes très particuliers d'usage). Pour ces raisons, et parce que les buts recherchés dans le projet Use-Me.Gov engageaient à trouver des moyens de favoriser l'efficacité d'interaction, nous avons choisi une autre approche basée sur la reconnaissance d'un sous-langage proche de la langue naturelle.

La langue naturelle possède en effet des propriétés qui la rendent avantageuse pour effectuer certains types de tâches. En particulier :

- la combinaison de plusieurs critères dans une même phrase revient à grouper plusieurs actions d'une interface tactile ; par exemple l'expression « du 20 mai au 30 mai dans le 18^{ème} et le 19^{ème} arrondissement » remplace la manipulation successive de plusieurs composants d'interaction ;
- l'emploi de quantifieurs et d'expressions logiques est pratique pour manipuler les expressions spatio-temporelles ; par exemple « tous les jeudi du mois de décembre » remplace avantageusement de multiples actions sur un calendrier ;
- la voix donne accès aux objets non visibles à l'écran et permet ainsi d'accéder à un ensemble d'informations bien plus vaste que ce qui peut être affichée sur un petit écran ; par exemple il n'est pas possible d'afficher un calendrier journalier avec les 12 mois de l'année tandis qu'il est tout à fait possible de désigner n'importe quel jour oralement ;
- la voix est potentiellement disponible à tout moment quelle que soit l'information affichée sur l'écran.

Le seul inconvénient majeur de la langue naturelle reste, encore aujourd'hui, le taux d'erreurs de reconnaissance. Toutefois, les erreurs étant fréquemment causées par la méconnaissance par l'utilisateur de ce qu'il peut dire et de comment il doit parler [5, 8], on peut faire l'hypothèse que l'utilisation d'un sous-langage dont les règles seraient aussi clairement indiquées que possible sur l'écran du dispositif mobile devraient fortement en diminuer l'occurrence. Notre choix s'est donc tourné vers la définition d'un tel sous-langage et la mise en œuvre de principes de conception pour en favoriser l'apprentissage rapide.

Ce sous-langage, dont la première version a été définie par le groupe de conception², est constitué d'un ensemble de phrases type dont certains tronçons sont optionnels et certains concepts associés à plusieurs mots (synonymes). Par exemple, pour effectuer une recherche d'information, la phrase type suivante a été définie : « Je voudrais des

² Dans le cadre d'un projet de développement industriel d'un service, cette première version du corpus aurait dû être complétée en s'appuyant sur des techniques de recueil impliquant des utilisateurs réels. Etant donné les objectifs de notre étude, cela n'était pas nécessaire.

informations sur /Canal d'Information/ dans /Arrondissement(s)/ pour /Date(s)/. » L'utilisateur n'est pas obligé de faire une phrase et peut formuler sa requête avec uniquement un, deux ou trois critères. Pour chaque valeur de critère, plusieurs mots ou expressions ont été prévus. La flexibilité offerte par ce type de sous-langage a l'intérêt de prendre en compte les attentes des utilisateurs à la fois novices mais aussi experts, lesquels peuvent par exemple formuler une demande uniquement avec une suite de mots clés.

Du point de vue de la conception, l'un des problèmes qui s'est posé était de rendre apparentes les phrases ou expressions attendues. Les sections suivantes décrivent les remises en cause dans la structuration de l'interface et la présentation de l'information qui nous a semblé nécessaire pour aller dans ce sens et tirer pleinement partie de la voix.

Mise à plat de l'arborescence

Deux facteurs liés à l'introduction de la voix en langue naturelle nous ont amené à réorganiser les pages pour parvenir au diagramme de la Figure 5. La commande vocale en langage naturel permet en théorie de gommer toute arborescence. Dans notre prototype, cela s'illustre par le fait qu'il devient possible d'effectuer une requête complète de recherche immédiate en une seule phrase (ex., « Je veux des informations sur la circulation aujourd'hui dans le 11^{ème} et le 12^{ème} arrondissement »). La voix rend aussi possible les requêtes simples d'abonnement et de désabonnement à tout moment. Dans ces conditions, nous avons décidé de regrouper les pages Accueil, Recherche et Abonnement sur une nouvelle page Accueil montrée sur la Figure 6.

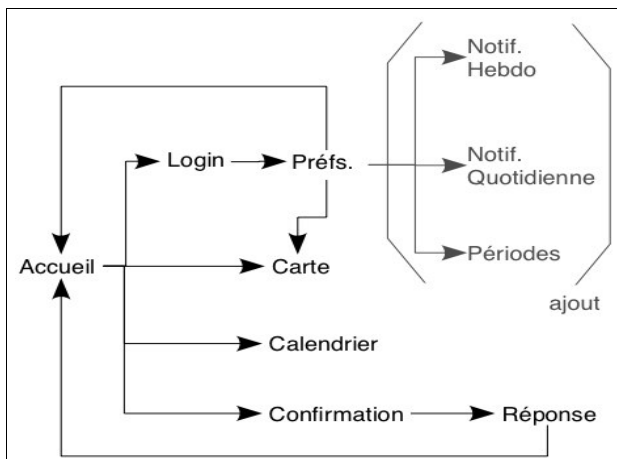


Figure 5: Nouveau diagramme de navigation entre pages dans la version multimodale

En revanche, il n'a pas été jugé souhaitable de regrouper la page Préférence avec les trois précédentes, même si en théorie l'utilisateur peut employer une combinaison multicritère pour modifier ses préférences. Les phrases correspondantes sont apparues complexes et, par conséquent, peu probables, comme l'illustre cet exemple : « je veux modifier mon abonnement au sport

en ajoutant le 2^{ème} arrondissement et en ne m'intéressant qu'aux événements du week-end ».



Figure 6: Nouvelle page d'accueil dans la version multimodale

Cette réorganisation permet d'anticiper des gains d'efficacité y compris avec l'usage de la modalité tactile. Ainsi, la profondeur maximale de l'interaction est maintenant de 1 page pour la recherche immédiate et de 4 pages au lieu de 5 pour la gestion des abonnements. Par ailleurs, l'entrée vocale est la modalité la plus efficace pour effectuer certaines requêtes. Ainsi une recherche immédiate avec les trois critères (canal, lieu et temps) nécessite dans le meilleur des cas avec la maquette tactile 9 désignations alors que trois désignations suffisent (Push-To-Talk, envoi de la requête, confirmation) avec une requête vocale.

Enrichissement des critères

Le fait que la voix autorise un plus grand espace de choix et un gain en nombre de pages peut permettre d'ajouter des pages pour enrichir les possibilités de sélection de certains critères et augmenter la redondance entre les deux modalités. Ces pages mentionnées en haut à droite sur la Figure 5 affinent la sélection du temps des événements ou de la notification dans les préférences.

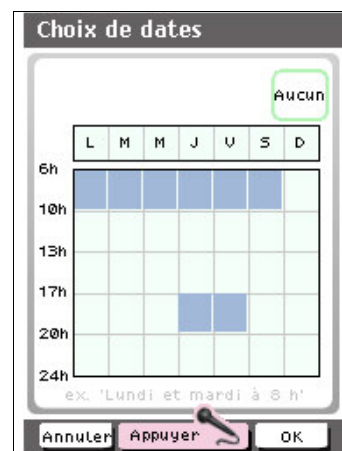


Figure 7: Nouvelle page de sélection des périodes intéressantes sur la semaine type avec sélections multiples

La Figure 7 montre ainsi la nouvelle page qui sert à définir la liste de créneaux horaires auxquels l'utilisateur s'intéresse sur une semaine type. Cette page permet d'effectuer un équivalent gestuel à des expressions tu type « tous les jeudi soirs ». La page de la Figure 7 remplace le menu déroulant de la Figure 2 qui ne propose que trois choix (semaine entière, jours ouvrés et week-end).

Redéfinition du guidage

L'introduction de la voix toute seule ne suffit pas à améliorer la facilité d'apprentissage pour deux raisons. D'une part, elle s'accompagne d'une réduction de la profondeur de l'arborescence des pages au profit de la largeur, ce qui peut constituer un facteur de surcharge de l'écran. D'autre part, le choix d'un sous-langage introduit le besoin de suggérer à l'utilisateur ce qu'il est en mesure de dire pour que la machine le comprenne. Pour répondre à ce besoin, la solution consiste à redéfinir le guidage.

Un nouveau guidage est indispensable tout d'abord pour informer l'utilisateur de la présence de la commande vocale. C'est le rôle du bouton avec un dessin de microphone et le label « appuyer/parler » des Figures 6 et 7. Ce bouton met en évidence le moyen de déclencher la reconnaissance vocale. Il possède deux états, encodés par des couleurs selon que la reconnaissance vocale est ouverte ou fermée.

Au cours de la re-conception, certaines fonctionnalités ont été jugées inadaptées à la voix, comme l'entrée du mot de passe pour des raisons de sécurité, ou les validations du type appui sur un bouton « ok » qui ne sont pas plus efficaces en vocal dès lors qu'il faut activer le microphone par une action de pointage. Si une page ne contient aucune fonctionnalité vocale, le bouton n'est alors pas apparent. Cette alternance attire l'attention et doit conduire à bien distinguer les fonctions possédant une entrée vocale des autres.

Le guidage est aussi nécessaire pour suggérer de manière transparente le vocabulaire et les types de phrases prononçables. Pour cela plusieurs indices visuels sont intégrés au graphisme. Premièrement, l'activation des mots pertinents dans la mémoire de l'utilisateur est favorisée par l'affichage des termes prononçables. Ces termes sont distingués des autres informations textuelles de la page par des guillemets (voir fig. 6). Deuxièmement, des exemples de requêtes sont affichés près des widgets associés aux commandes vocales. Ces exemples proposent différentes énonciations acceptables. La Figure 8 montre la page de sélection des arrondissements avec deux exemples de requêtes en grisé au-dessus de la carte.

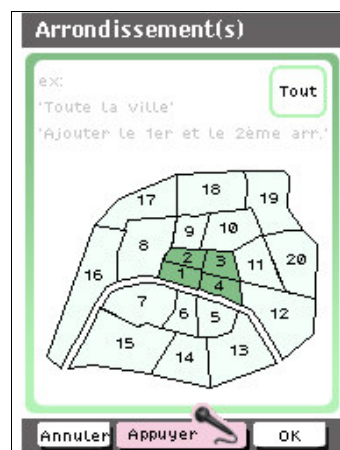


Figure 8: Page de sélection des arrondissements avec guidage vocal explicite dans la partie supérieure

Enfin, le guidage doit suggérer les possibilités de combinaison multicritères lorsqu'elles existent. Pour cela, l'interface est structurée sous formes de phrases naturelles : les labels et les champs participent activement à cette organisation comme le montre la page d'Accueil (fig. 6). Pour présenter l'ensemble des critères sur le même écran sans pour autant compromettre la lisibilité, des règles de structuration graphique telles que le regroupement par la couleur et l'utilisation de l'espacement ont été appliquées. Cette structuration en phrases a pour but d'amorcer chez l'utilisateur des structures de commandes vocales plus complexes et naturelles que de simples énonciations de mots clefs.

Détection des erreurs

Les retours d'information après chaque commande vocale jouent un rôle prépondérant dans la détection des erreurs. Ceux-ci sont effectués en deux temps : une animation qui attire l'attention vers les champs concernés suivie par la modification du contenu de ces champs.

L'animation consiste en l'apparition pendant quelques secondes d'un rectangle orange au dessus du champ qui vient d'être modifié par la commande. Cette apparition oriente le regard vers les zones pertinentes et minimise les chances de ne pas percevoir les conséquences d'une requête. La mise à jour du contenu des champs qui suit affiche ce que le système a compris ou bien une forme abrégée.

L'écran de gauche de la Figure 9 montre les rectangles oranges affichés après la requête « je veux des informations sur le sport dans tous les arrondissements le prochain week-end ». Ceux-ci montrent que les 3 critères de la requête vont être mis à jour. L'écran de droite montre la mise à jour des champs qui suit et qui permet à l'utilisateur de vérifier l'interprétation du système. Lorsqu'il n'y a pas assez de place pour tout montrer une forme abrégée est affichée. Le champ sert de toute façon à ouvrir la page de sélection du critère correspondant, l'utilisateur peut donc s'y rendre pour vérifier l'interprétation du système et la corriger. Ces mêmes

détails sont récapitulés sur la page Confirmation avant l'envoi de la requête.

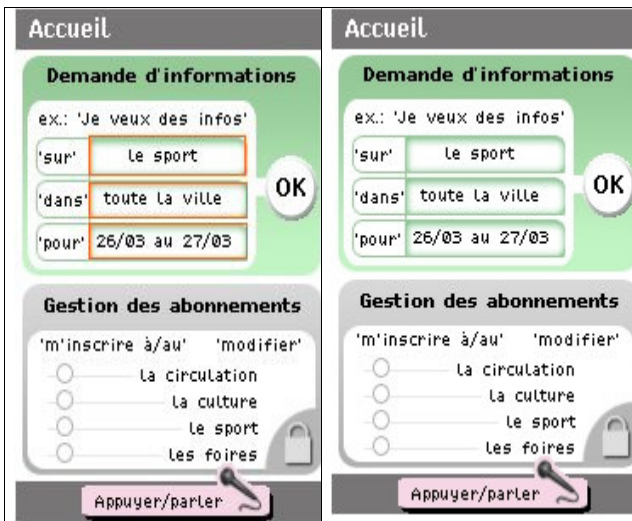


Figure 9: Retours d'information pour détecter les erreurs

Création de nouveaux composants d'interaction

Nous avons choisi de fournir une redondance complète entre les modalités tactiles et vocales pour ne pas privilégier une modalité au détriment de l'autre. Un autre avantage de cette redondance est qu'elle favorise les transferts d'apprentissage si l'utilisateur veut, par exemple, essayer la modalité vocale alors qu'il connaît déjà la modalité tactile. Cette recherche de redondance a exigé de créer de nouveaux composants d'interaction en remplacement de ceux qui existaient.

Ainsi, avec la voix, les critères spatio-temporels sont souvent groupés. Par exemple pour désigner des plages de temps, on dit naturellement : « du 19 au 31 janvier » ou, pour désigner une liste d'arrondissements, on peut dire : « du 1^{er} au 6^{ème} arrondissement ». Dans la maquette initiale, il n'existait pas de widgets permettant d'effectuer l'équivalent de ces commandes avec un seul geste. Ainsi en reprenant la Figure 1, la sélection d'une plage de temps nécessite quatre sélections au doigt et deux allers-retours avec la page de calendrier.

Nous avons donc créé des composants de sélection des plages temporelles basés sur des gestes continus pour sélectionner des plages en glissant le doigt ou le stylet sur plusieurs cases, comme avec le composant de la Figure 7. Le calendrier de la Figure 10 étend le même principe à un calendrier annuel : les semaines défilent automatiquement lorsque le geste s'immobilise au-dessus des zones de défilement.



Figure 10 : Calendrier remplaçant celui de la Figure 1, avec sélection multiples de plages de jours et les zones inférieure et supérieure de défilement.

Des techniques d'interaction du même ordre sont connues pour leur capacité à créer des actions continues qui enchaînent la sélection d'une commande et de plusieurs paramètres dans un même geste [4]. C'est l'application de l'idée suivant laquelle toute transaction qui peut être décrite par une seule phrase devrait pouvoir être exécuté par un seul geste [1].

Contextualisation des composants d'interaction

Les composants traditionnels conservés dans la maquette multimodale ont également subi des modifications liées à l'introduction de la voix. Ainsi, il est apparu nécessaire de rendre certains labels dynamiques pour qu'ils s'adaptent à l'état de l'interface et des retours d'information affichés après la dernière action tactile ou vocale. Par exemple, le label « pour » de la page d'Accueil précède le champ qui donne un retour d'information sur la période de temps sélectionnée. Son texte peut changer pour devenir « du » si le champ suivant contient une période comme « 5 - 10 mai » plutôt que « demain ». Dans le même ordre d'idée, le retour « 5 - 10 mai » est également devenu « 5 au 10 mai ». L'objectif est ici de maintenir le guidage vocal opérationnel quel que soit le dernier état atteint.

DISCUSSION

Dans l'expérience rapportée ici, l'introduction de la voix en langage naturel a imposé des modifications profondes de l'interface initialement monomodale. La reconception a procédé moins par adaptations mineures de l'existant que par une redéfinition assez globale de l'interaction.

Plusieurs arguments nous font penser que cela n'est pas lié à des particularités de notre projet. Il semble en effet plus raisonnable de penser qu'une telle redéfinition est inévitable, même si elle n'atteindra pas toujours le même degré, car elle semble être la conséquence de certaines propriétés de l'interaction vocale en langage naturel. Citons en trois majeures :

- le langage naturel tend à faire disparaître la notion d'arborescence si structurante pour l'interaction graphique ;
- l'interaction vocale suppose de redéfinir le guidage pour rendre visible ce qu'on peut dire et favoriser ainsi un usage aussi immédiat que possible ;
- l'interaction vocale conduit, dans un effort de cohérence avec les autres modalités disponibles, à redéfinir certains composants d'interaction existants.

Partant de là, nous pouvons tirer deux enseignements à l'attention des concepteurs d'interfaces multimodales :

1. Lors de la conception d'un (nouveau) service, l'idéal serait d'*anticiper l'évolution vers le multimodal* et de concevoir dès le départ l'interface dans cette optique. Cela impliquerait notamment d'entamer la conception en partant des buts de l'utilisateur, en partie atteignable par un seul acte de communication, et non d'une décomposition en actions élémentaires. L'un des éléments structurant majeur dans ce processus est alors la recherche d'un guidage efficace sur ce que l'on peut dire. On peut noter que dans un domaine connexe, celui des services multi-support (PC, PDA, Téléphone, ...), certains auteurs prônent de la même façon de commencer par produire une conception en envisageant le support le plus riche en entrée-sortie et de dégrader ensuite la présentation d'information pour l'adapter aux autres supports [3].
2. Bien évidemment, l'anticipation de l'environnement multimodal exact n'est pas toujours possible à un temps *t* ou fait face à de nombreuses incertitudes (quand ? quelle modalité sera ajoutée ? etc.). Dans ce cas, la leçon à tirer de notre expérience serait de *ne pas sous-estimer le temps de reconception* du service lors de l'évolution vers une interface multimodale, ce qui est tentant si l'on envisage l'ajout d'une modalité uniquement comme une adaptation de l'existant. Dans notre cas, ce temps a été quasiment équivalent au temps mis pour la réalisation de la première maquette.

REMERCIEMENTS

Nous tenons à remercier toute l'équipe du projet Use-Me.Gov et plus particulièrement les membres d'INDRA

et de FIT. Nous remercions aussi Aurélie Chavanne pour sa contribution à la conception graphique du prototype.

BIBLIOGRAPHIE

1. Buxton, W. *Chunking and Phrasing and the Design of Human-Computer Dialogues*. In Proceedings of the IFIP World Computer Congress, Dublin, Ireland, 1986, pp. 475-480.
2. Chatty, S., Sire, S. and Lemort, A. *Vers des outils pour les équipes de conception d'interfaces post-WIMP*. In Actes d'IHM'04, Namur, Belgique, ACM Press, 2004, pp. 45-52.
3. Florins M. & Vanderdonck (2004) Graceful degradation of user interfaces as a design method for multiplatform systems. *Proceedings of the 9th International Conference on Intelligent User Interface*, Funchal, Madeira, Portugal, ACM Press, 2004, pp. 140-147.
4. Guimbretière, F. and Winograd, T. *FlowMenu: combining command text and parameter entry*. In *Proceedings of UIST'2000*, ACM Press, 2000, pp. 213-216.
5. Karsenty, L. *Shifting the design philosophy of spoken natural language dialogue: From invisible to transparent systems*. *International Journal of Speech Technology*, 5(2), 2002, pp. 147-158.
6. Nigay, L. and Coutaz, J. *Espaces de conception des interfaces multimédia et multimodales*. *Techniques et Sciences de l'Informatique*, 15(9), 1996, pp. 1195-1225.
7. Schlienger, C., Valès, S. and Chatty, S. *Une expérience de conception et de prototypage d'interfaces*. In Actes d'IHM'04, Namur, Belgique, ACM Press, 2004, pp. 165-172.
8. Yankelovich, N., Levow, G. A. and Marx, M. *Designing SpeechActs: Issues in Speech User Interfaces*. In Proceedings of CHI '95, Denver, CO, May 7-11, 1995.
9. Wharton, C., Rieman, J., Lewis, C. and Polson, P. *The cognitive walkthrough method: A practitioner's guide*. In Nielsen, J., & Mack, R. L. (Eds.), *Usability inspection methods*. NY: John Wiley & Sons, 1994, pp.105-140.