Informations Personnelles	Waseem SAFI
	401 Grand Parc - App. 1226 - 14200 Hérouville Saint Clair
	Portable: +33 (0)6 84 93 73 46, Téléphone: +33 (0)2 31 56 73 98
	waseem.safi@unicaen.fr, waseemm2005@yahoo.co.nz
	https://www.greyc.fr/users/safiws
	Sexe: M
	Date de naissance: 19/10/1981 | Nationalité: Syrien

Diplômes	
	2013- 2016 : Doctorat / Bourse d'université de Caen Normandie – Caen/France Laboratoire : GREYC (Groupe de Recherche en Informatique, Image, Automatique et Instrumentation) Équipe : HULTECH (HUman Language TECHnology)
	2009-2010 : M2Recherche Informatique / université Paris-Sud - Paris/France (Bourse de l'institut Supérieur de Sciences Appliquées et Technologie / Damas - Syrie)
	2008-2009 : M2Pro VCIEL [visualisation et conception infographique en ligne] / Lyon2-France (Bourse de l'ambassade française en Syrie) – Master à distance.
	2008-2010 : M2 Complémentaire E-Services / FUNDP Namur - Belgique (Bourse de l'Agence Universitaire de la Francophonie (AUF) à Damas - Syrie) – Master à distance.
	2000-2005 : Génie informatique (Bac+5) / université de Damas_ Syrie (spécialisation : l'Intelligence Artificielle)
	2000 : Baccalauréat scientifique

Expérience de recherche et d'enseignement	
	2016 – 2017 : ATER - Attaché Temporaire d'Enseignement et de Recherche (TP-TD niveaux L1 - L2 - L3- M1) / 192H / département d'informatique / Université de Caen Normandie / Caen / France
	2013- 2016 : Doctorant / université de Caen Normandie – Caen/France Laboratoire : GREYC (Groupe de Recherche en Informatique, Image, Automatique et Instrumentation) Équipe : HULTECH (HUman Language TECHnology)
	2014-2015 : Enseignement vacataire (TP niveaux L3- M1) – 64H-

	département d'informatique / Université de Caen Normandie / Caen / France **2013-2014 :** Enseignement vacataire (TP niveaux L3- M1) – 64H- département d'informatique / Université de Caen Normandie / Caen / France **2010-2013 :** Enseignement (TP-TD niveaux M1-M2) / l'institut Supérieur de Sciences Appliquées et Technologie (Damas - Syrie), - Travailler en tant qu'ingénieur informatique à l'institut Supérieur de Sciences Appliquées et Technologie (Damas - Syrie), - Encadrant des projets (étudiants M1 et M2) **2010 :** Stage M2 Recherche – laboratoire LIMSI - université Paris-Sud - Paris/France **2006-2009 :** Enseignement (TP-TD niveaux M1-M2) / l'institut Supérieur de Sciences Appliquées et Technologie (Damas - Syrie), - Encadrant des projets (étudiants M1 et M2)
Expérience professionnelle	**2005-2006 :** Entreprise de Télécommunication (Syriatel – Damas/Syrie) / Java programmeur + bases de données Oracle **2000-2005 :** Entreprise des services du Web (www.vermeg.com - Damas/Syrie) / Java programmeur + bases de données

Expérience professionnelle des langages de programmation	Java, PHP, C++, C#, Python Java: J2ME, J2SE, J2EE, JSP- Servlets – JMF - EJB- J2D – J3D- OGRE4J, JESS (Java Expert System). JDBC connecteurs: MS Access –MySQL- Oracle- SQL Server, Web serveurs: Tomcat, WebSphere, WebLogic. Android Mobile: Eclipse ADT, Android studio.
Expérience professionnelle des bases de données	Oracle9i: (SQL, PLSQL, FUNDAMENTAL1-2, Forms), MySQL, SQL Server, SQLite.
Expérience professionnelle des langages graphiques	Blender 3D Animation, OpenWonderland VW, VRML, J3D

Langue maternelle	Arabe
Langues parlées et écrites	Française, et Anglaise

2016

- Safi, W., Maurel, F., Routoure, J-M., Beust, P., Molina, & M., Sann, C. (2016). "An Empirical Study for Examining the Performance of Visually Impaired People in Discriminating Ranges of Frequencies through Vibro-tactile Feedbacks". The International Workshop on the Future of Books and Reading in Human-Computer Interaction associated to NordiCHI´16, the 9th Nordic Conference on Human-Computer Interaction, October 23-27, 2016, Gothenburg, Sweden.

2015

- Safi, W., Maurel, F., Routoure, J-M., Beust, P., & Dias, G. (2015). "Blind Navigation of Web Pages on Touch Screen Devices". First Workshop ITI - Interface Tal-Ihm : Handicap – Texte Interprétation, associated with 22ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN2015). Caen, France, June 22.
- Safi, W., Maurel, F., Routoure, J-M., Beust, P., & Dias, G. (2015). "Web-Adapted Supervised Segmentation to Improve a New Tactile Vision Sensory Substitution (TVSS) Technology". 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015). London, UK, June 2-5.
- Safi, W., Maurel, F., Routoure, J-M., Beust, P., & Dias, G. (2015). "Supervised Segmentation of Web Pages for Vibro-Tactile Access on Touch-Screen Devices". 1st International Symposium on Web Algorithms (ISWAG 2015). Deauville, France, June 2-3.

2014

- Safi, W., Maurel, F., Routoure, J-M., Beust, P. & Dias, G. (2014). "Blind Browsing on Hand-Held Devices: Touching the Web to Understand it Better". Data Visualization Workshop (DataWiz 2014) associated to 25th ACM Conference on Hypertext and Social Media (HYPERTEXT 2014 – Rank A). Santiago, Chile, September 1.
- Safi, W., Maurel, F., Routoure, J-M., Beust, P. & Dias, G. (2014). "A Hybrid Segmentation of Web Pages for Vibro-Tactile Access on Touch-Screen Devices". 3rd Workshop on Vision and Language (VL 2014) associated to 25th International Conference on Computational Linguistics (COLING 2014 – Rank A). Dublin, Ireland, August 23.

2013

- Fabrice Maurel, Waseem Safi, Pierre Beust, Jean-Marc Routoure, "Navigation aveugle sur dispositifs mobiles : toucher le Web... pour mieux l'entendre", 16 Colloque International sur le Document Electronique (CIDE 2013), Lille, France, November.

2012

- Safi, W., AL Dakkak, O., and AL Aboud, W., "Designing a Question Analysis Component and Indexing Arabic Corpora in An Information Retrieval Systems for Question-Answering". 8th International Computing Conference in Arabic: ICCA 2012, Cairo, Egypt, 2012, December.
- Safi, W., Raad, H., and AL Dakkak, O., and "Designing a Three-Dimensional model of the human Face and deploying it on the Web in order to simulate the Speech Pronounciation and Emotion Representation". 8th International Computing Conference in Arabic: ICCA 2012, Cairo, Egypt, 2012, December.
- Safi, W., AL Dakkak, O., and Kadoura, E., "Designing a Three-Dimensional model of the human body and deploying it on the Web in order to simulate the movements of sign language Arabic". 8th International Computing Conference in Arabic: ICCA 2012, Cairo, Egypt, 2012, December.

2009

- Ghneim, N., Safi, W., and Al Said Ali, M., “Building a Framework for Arabic Ontology Learning”. The 13th IBIMA conference on Knowledge Management and Innovation in Advancing Economies, Marrakech, Morocco, 2009, November.
- Ghneim, N., E., Karhely, and Safi, W., “First Step of Building an Arabic FrameNet (AFN)”. The 13th IBIMA conference on Knowledge Management and Innovation in Advancing Economies, Marrakech, Morocco, 2009, November.

Activités d'enseignement :

2016 – 2017 : ATER - Attaché Temporaire d'Enseignement et de Recherche (TP-TD niveaux L1 - L2 - L3- M1) / 192H / département d'informatique / Université de Caen Normandie / Caen / France

Composante d'enseignement	Formation	Période	Enseignement	TP/TD
UFR des Sciences	Licence 1A Informatique	S1	ULI1G - Méthodologie	33
UFR des Sciences	Licence 2A Informatique	S1	ECI32 - Programmation orientée objet	16,5
UFR des Sciences	Licence 2A Informatique	S1	ECI33 - Documents structurés, architecture orientée Services	24
UFR des Sciences	Licence 2A Informatique	S1	ECI35 - Traitement automatique des langues	16,5
UFR des Sciences	Licence 2A Informatique	S2	ECI44 - Programmation fonctionnelle	24
UFR des Sciences	Licence 3A Informatique	S2	ECI66 - Créativité logicielle et nouvelles interfaces	48
UFR des Sciences	M1 Informatique	S2	UMI2D - Technologies (Java Enterprise Edition)	20
UFR des Sciences	M1 Informatique	S2	UMI2N - Moteurs de recherche	10
Total des heures de service :				192

2014-2015 : Enseignement vacataire (TP niveaux L3- M1) / 64H / département d'informatique / Université de Caen Normandie / Caen / France

Composante d'enseignement	Formation	Période	Enseignement	TP
UFR de Sciences	Licence 3A Informatique	S2	ECI66 - Créativité logicielle et nouvelles interfaces	48
UFR de Sciences	M1 Informatique	S1	UMI1G - Technologies XML	6
UFR de Sciences	M1 Informatique	S1	UMI1J - Ingénierie des langues	10
Total des heures de service :				64

2013-2014 : Enseignement vacataire (TP niveaux L3- M1) – 64H- département d'informatique / Université de Caen Normandie / Caen / France

Composante d'enseignement	Formation	Période	Enseignement	TP
UFR de Sciences	Licence 3A Informatique	S2	ECI66 - Créativité logicielle et nouvelles interfaces	48
UFR de Sciences	M1 Informatique	S1	UMI1G - Technologies XML	6
UFR de Sciences	M1 Informatique	S1	UMI1J - Ingénierie des langues	10
Total des heures de service :				64

2010-2013 : Enseignement (TP-TD niveaux 4ème et 5ème années) / l'institut Supérieur de Sciences Appliquées et Technologie (ISSAT) à Damas/Syrie

Heures de service / an

Composante d'enseignement	Formation	Période	Enseignement	TP/TD
ISSAT	4ème année informatique	S1	Programmation orientée objet (Java)	33
ISSAT	4ème année informatique	S1	Séminaires (3 groupes)	30
ISSAT	5ème année informatique	S1	Traitement automatique des langues	16.5
ISSAT	4ème année informatique	S2	Recherche opérationnelle	16.5
ISSAT	4ème année informatique	S2	Algorithmes avancés	16.5
Total des heures de service / an :				112,5

2006-2009 : Enseignement (TP-TD niveaux 4ème et 5ème années) / l'institut Supérieur de Sciences Appliquées et Technologie (ISSAT) à Damas/Syrie

Heures de service / an

Composante d'enseignement	Formation	Période	Enseignement	TP/TD
ISSAT	4ème année informatique	S1	Programmation orientée objet (Java)	33
ISSAT	4ème année informatique	S1	Séminaires (3 groupes)	30
ISSAT	5ème année informatique	S1	Traitement automatique des langues	16.5
ISSAT	4ème année informatique	S2	Recherche opérationnelle	16.5
ISSAT	4ème année informatique	S2	Algorithmes avancés	16.5
Total des heures de service / an:				112,5

Activités de recherche :
2013- 2016 :

Doctorant : université de Caen Normandie – Caen/France
laboratoire : GREYC (Groupe de Recherche en Informatique, Image, Automatique et Instrumentation)
équipe : HULTECH (HUman Language TECHnology)
sujet de thèse : Non-visual Vibro-tactile Navigation of Web Pages on Touch Screen Devices
directeur de thèse : Pierre BEUST, Co-encadrement : Fabrice MAUREL, université de Caen Normandie
domaine de recherche : IHM Interface Homme Machine
objectif général de la thèse : étudier l'effet de la modalité vibrotactile sur la capacité des personnes non voyantes à percevoir la structure visuelle (mise en forme) des pages Web.
problématique : Comment concevoir des systèmes interactifs favorisant l'exploitation non visuelle des processus cognitifs (skimming et scanning) pour faire émerger des stratégies de lecture non-visuelle efficace ?
résumé de la thèse : Les personnes voyantes interprètent visuellement une grande quantité des informations visuelles et comprennent des mises en forme complexes de pages web en quelques secondes. Tout d'abord, ils naviguent une page web rapidement pour obtenir un aperçu global de la structure du contenu. Ce processus est appelé skimming. Après, ils lisent le contenu des parties intéressantes en détails par suivant plusieurs méthodes de lecture. Ce processus est appelé scanning. Ce type de navigation des informations permet au lecteur de percevoir rapidement la mise en forme du document et sa structure. En outre, il oriente rapidement l'attention du lecteur sur les parties intéressantes de manière systématique et prévisible. Malheureusement, les personnes ayant une déficience visuelle ne peuvent pas faire cette interprétation dans la même efficacité. Ils perdent beaucoup de temps pour explorer l'architecture des pages web en utilisant les logiciels informatiques actuels spécialisés pour les personnes ayant une déficience visuelle. Ces logiciels permettent de naviguer séquentiellement les pages web, une colonne à la fois, une ligne à la fois, et un mot à la fois. Ils ne parviennent pas à transmettre l'ordre naturel des éléments sur une page web.

L'interprétation de la disposition spatiale d'une page web est souvent indispensable pour comprendre son contenu, et il permet à l'utilisateur de traiter des informations complexes en les réduisant à un nombre gérable d'unités. La mise en forme, et les indices spatiaux des pages web sont très importants pour permettre nombreuses tâches, et pour guider l'utilisateur à analyser et à trouver des éléments de données, et pour réaliser de tâches plus compliquées tels stocker des informations, trouver rapidement les informations pertinentes, pointant directement toute information présente dans une page web.

Ce travail présente une approche pour un accès non-visuel aux pages web. L'approche proposée est basée sur une modalité vibro-tactile pour remplacer le processus de survol lors de la navigation des pages web affichées sur des appareils mobiles avec écrans tactiles. L'approche aide les personnes ayant une déficience visuelle à obtenir un aperçu global de la structure du contenu, et de percevoir rapidement la mise en forme de la page web.

Notre contribution est la conception d'un système de substitution sensorielle de la vision tactile original. Ce système convertit les structures visuelles qui représentent la mise en forme d'une page web à des vibrations tactiles. Cette approche de navigation reproduit l'équivalent d'une exploration visuelle classique d'un document sur la base d'une vibration de luminosité. Cette approche proposée pourrait être considérée comme une nouvelle solution de navigation non-visuelle pour l'exploitation des informations spatiales des pages web.

Des séries d'expériences ont été menées avec des personnes ayant une déficience visuelle a confirmé l'hypothèse que les personnes ayant une déficience visuelle peuvent explorer des formes géométriques graphiques présentés sur un appareil mobile avec un écran tactile, et ils peuvent percevoir leurs variétés de la taille, la forme, la topologie, les relations spatiales, et les contenus sémantiques en utilisant des vibrations tactiles.

Résultats :

- la modalité vibrotactile avec la TactiNET peut favoriser l'exploration de formes géométriques présentées sur dispositif mobile à écran tactile.
- la structure visuelle peut être traitée comme une configuration de formes géométriques de base.
- de nombreux éléments pour l'automatisation de cette transformation et son évaluation ont été proposés et mis en œuvre :
    - algorithmes de traitement de pages Web : extraction, segmentation, clustering, génération de formes graphiques et tactiles,
    - dispositif hardware et software TactiNET,
    - plateforme d'expérimentation Androïd simple et riche.
- la démarche expérimentale a permis :
    - de mieux comprendre la façon dont les personnes non-voyantes appréhendent les pages web
    - de préciser les bases d'un langage graphique et vibrotactile

2010-2013 :

Encadrement des projets d'étudiants à l'institut Supérieur de Sciences Appliquées et Technologie (ISSAT) et à université de Damas (Damas - Syrie) :

- concevoir un modèle tridimensionnel du corps humain et le déployer sur le Web pour simuler les mouvements de la langue des signes arabe : ISSAT, 4ème année informatique, 2012.
- concevoir un composant d'analyse des questions et indexer des corpus en langue arabe dans un système de récupération d'information pour répondre aux questions : ISSAT, 4ème année informatique, 2012.
- concevoir un modèle tridimensionnel du visage humain et le déployer sur le Web afin de simuler la prononciation du discours et la représentation émotionnelle : ISSAT, 4ème année informatique, 2012.
- construire un modèle 3D pour représenter les mouvements du visage pendant la parole et les mouvements qui représentent la langue des signes arabe : ISSAT, 5ème année informatique, 2011.
- construire un système de réponse aux questions : université de Damas, département d'informatique, 5ème année informatique, 2010.

2010 (du 1 avril au 31 août 2010):

stage m2rech informatique : Sélection de passages pour répondre à des questions complexes
laboratoire : LIMSI-CNRS Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur / université Paris Sud / Paris / France
encadrement : Véronique MORICEAU et Xavier TANNIER
Résumé : Le système FIDJI est un système de question-réponse en domaine ouvert pour le français et l'anglais, qui utilise l'analyseur syntaxique XIP, et le moteur de recherche Lucene, avec d'autres techniques dans le but de répondre à différents types de questions.

FIDJI analyse syntaxiquement la question et extrait les informations utiles tel que le type de la question, le type attendu de la réponse, les mots clés, et d'autres informations, cherche des documents correspondant aux mots clés de la question, analyse ces documents, et extrait les phrases qui conviennent à la question.

Un des types de questions auxquelles FIDJI répond est les questions complexes qui dans beaucoup de cas commencent par «comment» ou «pourquoi». Les réponses à ces questions sont des passages à l'inverse d'autres types des questions dont les réponses sont courtes telles que les questions factuelles ; le but du stage est de définir un ensemble de critères (relations syntaxiques, termes de la question, marqueurs discursifs, ...) permettant de décider si le passage sélectionné est de nature à répondre à une question complexe.

L'objectif du stage : est de définir un ensemble de critères qui aide à améliorer la méthode de recherche et d'extraire des passages qui contiennent les réponses aux questions complexes posées. Dans le but de réaliser cet objectif, le travail durant le stage a été organisé de la façon suivante:

- étudier l'état d'art des systèmes question-réponse globalement, et plus spécialement les systèmes question-réponse complexes,
- étudier les phénomènes linguistiques « relations de causalité » et « anaphore » et leurs effets sur les systèmes question-réponse,
- étudier l'architecture de FIDJI,
- analyser les questions complexes et leurs réponses dans les corpus Quaero09 et RespubliQA09,
- comprendre l'architecture de l'analyseur XIP globalement, la sortie de XIP et les règles d'analyse des questions et des textes,
- proposer des algorithmes pour atteindre l'objectif du stage,
- programmer les algorithmes proposés,
- tester les algorithmes et évaluer les résultats.

Résultats du stage :

Les algorithmes proposés ont été appliqués sur les corpus CLEF05, Quaero09, et ResPubliQA09. Quelques tests ont augmenté le nombre des bonnes réponses, d'autres tests ont augmenté le nombre de documents récupérés par le moteur de recherche Lucene et qui contiennent des bonnes réponses. Par contre, dans quelques cas, certains tests ont diminué la performance du système, et dans d'autres cas, ils ont diminué le nombre de documents récupérés par Lucene et qui contiennent des bonnes réponses.

La plupart des améliorations sont sur le corpus ResPubliQA09, il s'agit en effet du corpus qui contient la plus grande partie des questions complexes, 49 % de ses questions sont de type Complexe, tandis que seulement 13.21 % des questions du corpus Quaero09 sont de type Complexe.Les algorithmes appliqués sur CLEF05 n'améliorent jamais la performance du système, ni pour augmenter le nombre de documents récupérés, ni pour faire l'extraction des bonnes réponses; dans le meilleur cas la performance diminue environ 6.5 % pour les bonnes réponses en premier position, c'est parce que CLEF05 ne contient pas de questions complexes, et la majorité sont questions factuelles.

Concernant le corpus Quaero09, les algorithmes ont augmenté le nombre de documents récupérés qui contient les bonnes réponses environ 1 %, mais aucun des algorithmes appliqués a amélioré l'extraction des bonnes réponses.

2007-2009 :

Encadrement des projets d'étudiants à l'institut Supérieur de Sciences Appliquées et Technologie (ISSAT) et à université de Damas (Damas - Syrie) :

- correction orthographique automatique pour les textes arabes : ISSAT, 5ème année informatique, 2009.
- construire une nouvelle hiérarchie pour les noms et verbes arabes et le déployer sur le Web (VerbNet): ISSAT, 4ème année informatique, 2008.
- Construction d'une nouvelle hiérarchie pour les concepts et relations arabes (FrameNet) : ISSAT, 4ème année informatique, 2008.
- Construction d'une nouvelle hiérarchie pour les noms, verbes et concepts arabes (Ontologie) : ISSAT, 4ème année informatique, 2008.
- moteur de recherche basé sur WordNet : université de Damas, département d'informatique, 4ème année informatique, 2008.
- un système d'étiquetage morpho syntaxique : ISSAT, 4ème année informatique, 2008.
- un système automatisé d'évaluer des essais en arabe : université de Damas, département d'informatique, 5ème année informatique, 2007.

Résumé

Les personnes voyantes interprètent visuellement une grande quantité des informations visuelles et comprennent des mises en forme complexes de pages web en quelques secondes. Tout d'abord, ils naviguent une page web rapidement pour obtenir un aperçu global de la structure du contenu. Ce processus est appelé Skimming. Après, ils lisent le contenu des parties intéressantes en détails par suivant plusieurs méthodes de lecture. Ce processus est appelé Scanning. Ce type de navigation des informations permet au lecteur de percevoir rapidement la mise en forme du document et sa structure. En outre, il oriente rapidement l'attention du lecteur sur les parties intéressantes de manière systématique et prévisible. Malheureusement, les personnes ayant une déficience visuelle ne peuvent pas faire cette interprétation dans la même efficacité. Ils perdent beaucoup de temps pour explorer l'architecture des pages web en utilisant les logiciels informatiques actuels spécialisés pour les personnes ayant une déficience visuelle. Ces logiciels permettent de naviguer séquentiellement les pages web, une colonne à la fois, une ligne à la fois, et un mot à la fois. Ils ne parviennent pas à transmettre l'ordre naturel des éléments sur une page web.

L'interprétation de la disposition spatiale d'une page web est souvent indispensable pour comprendre son contenu, et il permet à l'utilisateur de traiter des informations complexes en les réduisant à un nombre gérable d'unités. La mise en forme, et les indices spatiaux des pages web sont très importants pour permettre nombreuses tâches, et pour guider l'utilisateur à analyser et à trouver des éléments de données, et pour réaliser de tâches plus compliquées tels stocker des informations, trouver rapidement les informations pertinentes, pointant directement toute information présente dans une page web.

Ce travail présente une approche pour un accès non-visuel aux pages web. L'approche proposée est basée sur une modalité vibro-tactile pour remplacer le processus de survol lors de la navigation des pages web affichées sur des appareils mobiles avec écrans tactiles. L'approche aide les personnes ayant une déficience visuelle à obtenir un aperçu global de la structure du contenu, et de percevoir rapidement la mise en forme de la page web.

Notre contribution est la conception d'un système de substitution sensorielle de la vision tactile original. Ce système convertit les structures visuelles qui représentent la mise en forme d'une page web à des vibrations tactiles. Cette approche de navigation reproduit l'équivalent d'une exploration visuelle classique d'un document sur la base d'une vibration de luminosité. Cette approche proposée pourrait être considérée comme une nouvelle solution de navigation non-visuelle pour l'exploitation des informations spatiales des pages web.

Des séries d'expériences ont été menées avec des personnes ayant une déficience visuelle a confirmé l'hypothèse que les personnes ayant une déficience visuelle peuvent explorer des formes géométriques graphiques présentés sur un appareil mobile avec un écran tactile, et ils peuvent percevoir leurs variétés de la taille, la forme, les relations spatiales, et les contenus sémantiques en utilisant des vibrations tactiles.

Mots clés : Une déficience visuelle, Un accès non-visuel aux pages web, Informations visuelles, L'architecture des pages web, La structure du contenu, Une modalité vibro-tactile, Balayage, Appareils mobiles avec écrans tactiles.

Caen, le 31 janvier 2017

Pierre BEUST
Maître de conférences HDR en Informatique
GREYC CNRS-UMR 6072 Équipe Hultech
Université de Caen Normandie
Campus 2, Bd Mal Juin
14032 Caen Cedex
tel : 02 31 56 73 97
e-mail : pierre.beust@unicaen.fr

A qui de droit.

Objet : Recommandation de Waseem SAFI
pour le prix de thèse IFRATH-KAELIS 2016

Chers collègues,

Depuis juin 2013, j'ai eu l'occasion de travailler avec Waseem SAFI en dirigeant sa thèse avec Fabrice Maurel, maître de conférence en Informatique au laboratoire GREYC CNRS UMR 6072, comme co-encadrant. Durant ces 3 années, j'ai pu suivre de près son parcours et ses évolutions jusqu'à l'écriture de son mémoire (213 pages, 156 références bibliographiques) et la soutenance de son doctorat le 14 décembre 2016.

Le sujet de thèse de Waseem SAFI s'est inscrit dans le cadre d'un projet ANR, le projet ARTADN (Accès par Retour Tactilo-oral Aux Documents Numériques), coordonné par Fabrice Maurel ayant pour objectif de développer et de tester un dispositif d'accès vibro-tactile aux pages web pour des utilisateurs non voyants (la TACTINET).

Le sujet de thèse de Waseem SAFI s'inscrit dans le cadre de l'« e-accessibilité » appliqué aux personnes non voyantes désirant accéder à la structure globale des pages web. L'objectif est de concevoir un système qui convertit les structures visuelles d'une page web en des vibrations tactiles (un système tactile de substitution sensorielle de la vision), d'élaborer une plateforme expérimentale, et enfin d'évaluer expérimentalement différentes hypothèses.

Le travail demandé dans cette thèse était principalement une recherche en Informatique (plus précisément en Interactions Homme-Machine) mais aussi un travail réellement pluridisciplinaire allant de l'informatique et l'électronique jusqu'à la psychologie et les sciences cognitives : réalisation du dispositif électronique autonome intégrant des vibreurs haptiques et une liaison Bluetooth, programmation de l'interface sur tablette Android pour mise en place de tests psycho-sensoriels, conception des algorithmes de segmentation de pages web, réalisation de corpus de documents numériques, analyse des résultats scientifiques obtenus… Pour ces différentes tâches, monsieur Waseem SAFI a fait preuve d'autonomie et de rigueur. Il a parfaitement répondu aux attentes du projet.

Aujourd'hui les résultats sont très prometteurs. C'est un constat qui est ressorti de manière consensuelle du jury de soutenance. Les utilisateurs non voyants peuvent percevoir par la TACTINET les grandes structures visuelles et textuelles d'une page web auxquelles ils n'ont pas accès. Ils peuvent ainsi faire usage dans leur interaction avec la page web de "trottoirs textuels" reconnaissables par leurs profils vibratoires pour notamment mieux piloter une synthèse vocale à tel ou tel endroit d'une page web. Plusieurs pistes sont possibles dans la suite de la thèse de Waseem SAFI, notamment la possibilité d'établir un langage perçu par des senseurs vibratoires et/ou thermiques permettant de reconnaître la sémantique de tels ou tels composants d'une page web.

Dès que nous avons commencé à travailler ensemble, j'ai pu constater la grande motivation de Waseem SAFI pour la recherche en Informatique appliquée aux problèmes de handicap. Je souligne l'aisance technique de Waseem SAFI et son efficacité dans le développement logiciel et l'intégration de diverses technologies (notamment le développement pour terminaux mobiles sous Android). Il m'a souvent impressionné par sa grande capacité de travail et en sachant faire preuve d'une grande autonomie dans ses réalisations. Je retiens aussi que Waseem SAFI a su acquérir tout au long de ces dernières années une bonne culture pluridisciplinaire et mettre en œuvre une réelle démarche scientifique expérimentale. Sa capacité à multiplier les contextes d'expérimentation et ses propositions de protocoles avec des "vrais" utilisateurs le montrent bien. D'autre part, son activité de publication internationale est soutenue et fournie. Elle témoigne de ses évidentes qualités de chercheur.

Pour toutes ces raisons font que je recommande sans réserve la candidature Waseem SAFI au prix de thèse IFRATH-KAELIS 2016.

Pierre BEUST

Rapport sur le mémoire de thèse de Monsieur Wassem SAFI

Intitulé : «**Non-Visual Vibro-Tactile Navigation of Web Pages on Touche Screen Devices** »
Pour obtenir le grade de docteur en informatique et applications de l'Université de Caen - Normandie

Rapporteur : Khaldoun Zreik, Professeur des Universités, Université Paris 8

M. Waseem Safi a effectué ses travaux de recherche au laboratoire GREYC de l'université de Caen - Normandie, sous la direction de Monsieur Pierre Beust (MCF HDR), Monsieur Jean-Marc Routoure (Professeur) et Monsieur Fabrice Maurel (MCF).

Ce travail de recherche et de développement porte sur l'accès non visuel aux pages web. L'hypothèse centrale de ce travail de thèse suggère la modalité vibro-tactile pour naviguer et consulter des pages web (et documents numériques) affichées sur des appareils mobiles avec écrans tactiles. Cette approche a pour objectif de permettre aux personnes ayant de déficience visuelle d'obtenir un aperçu global, non visuel, de la structure de contenu d'une page web. Afin d'évaluer son hypothèse, M. Safi a conçu un système de substitution sensorielle de la vision tactile (TVSS) et a réalisé des expérimentations avec des personnes ayant de déficience visuelle et avec des personnes voyantes. Les résultats obtenus semblent, selon le candidat, consolider le fait que les personnes ayant une déficience visuelle peuvent explorer et percevoir, sur un appareil mobile avec un écran tactile, différentes formes géométriques et graphiques.

Le mémoire présenté par M. Safi est structuré en une introduction générale, 3 parties (regroupant 6 chapitres), des conclusions - perspectives, une liste de références bibliographiques et 4 annexes. L'ensemble présente 206 pages.

En **introduction** (2 pages), M. Safi rappel, brièvement, le contexte du travail de recherche, c'est à dire les problèmes et les limitations des solutions proposées permettant l'accès et la navigation des pages web à des personnes malvoyantes ou non voyantes, avant de présenter le plan du mémoire.

La première partie (56 pages) est intitulé « Targeted Users, State of the Art | Utilisateurs ciblés, état de l'art » présente l'état de l'art des offres technologiques s'adressant aux personnes malvoyantes ou non-voyantes. Cette partie comprend deux chapitres.
Le **premier chapitre** « Access to Spatial Information in Web Pages by Visually Impaired Persons | Accès à l'information spatiale dans les pages Web par des personnes malvoyantes » (17 pages) présente les concepts de l'accessibilité aux pages Web et les limites actuelles des technologies d'assistance dédiée aux personnes malvoyants (Braille, Synthèse de la Parole, OCR, écrans Tactiles). Ce chapitre se termine par la présentation de quelques concepts sur la perception de l'aménagement spatiale des pages Web, ainsi que certains principes de la modalité de perception vibro-tactile.
Le **deuxième chapitre** « Non-visual Access to Visual Information in Numerical Documents | Accès non visuel à l'information visuelle dans les documents numériques » (34 pages) définit des concepts de systèmes de substitution sensorielle de la vision tactile (TVSS). Trois catégories de systèmes de substitution sensorielle de vision tactile sont présentées (braille, vibration, spatiale). Pour terminer ce chapitre le candidat présente plusieurs systèmes de navigation, non visuelle, des pages web.

La deuxième partie du mémoire est intitulée « System Description, Hardware and Software implementation | Description du système, implémentation matérielle et logicielle » (72 pages). Cette partie présente l'objectif principal et l'hypothèse centrale du système proposé « TactiNET » (système sensoriel de visualisation tactile) ainsi que quelques détails techniques de cette solution. Ensuite, le candidat expose le schéma global du système et la méthodologie adoptée pour atteindre la solution proposée. Cette partie comprend 3 chapitres qui représentent les contributions principales, en matière de développement et d'expérimentation, de ce travail de thèse.
Le **troisième chapitre** « TactiNET Framework, Hardware Implementation | TactiNET, implémentation matérielle » (10 pages) présente les critères, les conditions et les propriétés souhaitées que doivent satisfaire les composants matériels (hardware prototype) implémentés dans TactiNET, ainsi que les deux dispositifs qui ont été conçus et produits au cours de ce travail (un prototype avec quelques limitations

concernant les actionneurs; un outil plus professionnel sans limitations concernant les actionneurs).
Le **quatrième chapitre** « Software Implementation | Implémentation logicielle » (28 pages) présente les principaux composants et ressources logiciels utilisés dans le développement du TactiNET. Le candidat y propose un algorithme de segmentation supervisée des pages Web dédié à l'accès vibro-tactile sur les dispositifs à écran tactile. Cet algorithme permet aux personnes malvoyantes de comprendre la page Web en 2 dimensions en transformant les pages Web en pages vibrantes à l'aide d'un langage graphique vibro-tactile (inspirée de la théorie de la Gestalt). Une comparaison entre la segmentation automatique et la segmentation manuelle des pages Web est présentée dans ce chapitre. De nombreuses expériences avec des personnes voyantes et non voyantes, utilisant TactiNET, ont été réalisées et exposées dans ce chapitre. La méthodologie de construction du corpus Web ART-ADN et d'extraction des représentants des pages web y est également présentée.
Le **cinquième chapitre** «Basic Parameters and Arguments of the Graphical Vibro-tactile language | Paramètres de base du langage graphique Vibro-tactile » (26 pages) présente trois expériences qui ont été menées pour concevoir le langage vibro-tactile graphique. La première expérience vise à examiner la performance des personnes malvoyantes dans la reconnaissance des formes grâce à des rétroactions vibro-tactiles. Les deuxième et troisième expériences visent à sélectionner les plages de fréquences et les valeurs d'amplitude à utiliser dans la génération des retours vibro-tactiles. Le candidat expose dans ce chapitre le protocole d'exécution de chaque expérience, les résultats obtenus et leurs analyses.

La troisième partie est intitulé « Recognizing the Structures of Web Pages Through Vibro-Tactile Feedbacks | Reconnaissance des structures des pages Web grâce aux retours vibrato-tactiles» (40 pages). Cette partie comprend un seul chapitre.
Le **cinquième chapitre** « Recognizing the Structures of Web Pages Through Vibro-Tactile Feedbacks | Reconnaissance des structures des pages Web grâce aux retours vibrato-tactiles » (38 pages), présente l'expérimentation principale qui a été menée auprès des personnes non-voyantes et malvoyantes afin d'examiner leurs capacités à reconnaître les structures de pages Web grâce à des retours vibro-tactiles. Dans ce chapitre le candidat introduit les objectifs de l'expérience menée, puis il présente les formes choisies pour représenter les structures de pages Web avant de présenter le protocole de l'expérience et les résultats qui sont discutés en fin de ce chapitre.

En « **Conclusion and Perspectives** | Conclusions et perspectives » (2 pages) le candidat dresse une courte synthèse des travaux réalisés en énonçant quelques perspectives d'évaluation et d'amélioration du système proposé.

Le mémoire se termine par:
- Une liste de **références bibliographiques** (4 pages)
- Quatre **annexes**
  - annexe A: An image processing based segmentation Approach | Une approche de segmentation basée sur le traitement d'image
  - annexe B: Manual and automatic web page segmentation comparison | Comparaison de segmentation de pages Web manuelle et automatique
  - annexe C: List of sites used to construct ART-ADN corpus | Liste des sites utilisés pour construire le corpus de ART-ADN
  - annexe D: Gray-scale images of representative web pages | Images au niveau des gris des pages Web représentatives

Remarques sur le fond
Le sujet de ce travail de recherche et de développement est de grand intérêt surtout en ce moment où l'organisation informationnelle mais aussi communicationnelle de la vie, aussi bien socioprofessionnelle que sociale, sont de plus en plus dépendantes des technologies de l'information et de la communication en générale et de la téléphonie intelligente « Smartphone » en particulière. Prendre en compte l'handicap pour éviter toute exclusion sociale et socioprofessionnelle est un objectif noble. On ne peut donc que saluer cette initiative entreprise par le laboratoire et par le candidat.

Certes, la problématique telle qu'elle est formulée, autour de la notion de l'écrémage des pages web, serait le point le plus original (pour non spécialistes des interfaces pour handicap) de ce travail. On apprécie l'important travail d'analyses et de veille technologique qui a permit la construction et l'élaboration de l'hypothèse centrale de ce travail. Cette approche de formulation de l'hypothèse semble plus pertinente que l'hypothèse retenue. Cette dernière serait plus techno-centrée que les analyses entreprises. Une fois problématiques et hypothèses sont arrêtées, la méthodologie de recherche et de développement qui a été adoptée semble très convenable.
M. Safi a opté pour un travail d'exploration, presque exhaustive, des offres technologiques issues des laboratoires de recherche et de l'industrie. Il a veillé à rappeler, soigneusement, les avantages et les limitations de chaque objet technologique exposé. Ceci dit, sa position par rapport à la technologie de Braille (codage Braille) reste discutable. Car la technologie vibro-tactile exigerait également une approche de codage à apprendre.
Les développements proposés ainsi que les expérimentations menées ont effectivement servit à renforcer l'hypothèse du travail, à savoir le choix de la modalité vibro-tactile. Mais cette validation, qui est certainement nécessaire, n'est pas complètement satisfaisante. Chose que le candidat observe directement ou indirectement tout au long du mémoire. Ce fait important, que ce travail a le mérite de faire émerger clairement, invite à repenser autrement la conception de l'interface et peut être de l'information. Aussi les résultats de ce travail devraient conduire à explorer différemment les champs d'exploration des capacités de perception et d'agir chez l'être humain, en situation d'handicap ou pas. Il en est de même en ce qui concerne la question d'apprentissage de codes (cas du Braille), où la latence, ou bien la difficulté d'apprentissage, ou bien la limitation d'usage, auraient du interpeller le candidat à réinterroger la philosophie de cette approche en tant que perspective.
On aurait aimé que ces derniers points soient développés ou bien évoqués dans le chapitre conclusions et perspectives.

<u>Remarques sur la forme</u>

Le mémoire est rédigé en anglais. La rédaction est claire et facile à lire (pour un lecteur dont l'anglais n'est pas la langue maternelle). La répartition en partie et en chapitre n'est pas assez équilibrée, surtout la dernière partie qui se limite à un seul chapitre. Sinon, on apprécie que chaque partie et chaque chapitre est précédé par une introduction et suivi par une conclusion. M. Safi ne propose qu'un seul résumé en langue française (1 page) en début du mémoire, cela aurait été plus intéressant d'associer un résumé en français au début de chaque partie ou chapitre. Les figures, tableaux et schémas sont de bonnes qualités. On apprécie la richesse de la liste des références bibliographiques dont la lecture aurait été plus facile en séparant les références « webographiques ». Par souci de clarté du propos, le mémoire observe quelques très légères répétitions (rappelant la problématique de la recherche). Finalement, on regrette que le chapitre consacré aux conclusions et perspectives soit très court.

<u>Conclusion du rapport</u>

L'intérêt du sujet abordé ainsi que l'importance et le volume du travail de recherche et de développement réalisé, me mènent à donner un avis favorable à ce que les travaux de M. Waseem Safi soient présentés devant le jury d'examen pour l'obtention du diplôme de docteur de l'Université de Caen-Normandie.

Fait à Paris le 17 novembre 2016

Pr. Khaldoun Zreik
Directeur du département des Humanités Numériques (UFR MITSIC) - Université Paris 8
Responsable de l'équipe CiTU-Paragraphe, Université Paris 8 | Université de Cergy Pontoise

Rapport sur le mémoire présenté par
Monsieur **Waseem SAFI**
en vue de la soutenance d'une thèse de doctorat
de l'**Université de Caen**.
Mention : Informatique

Etabli par **Mustapha Mojahid**
Maître de Conférences – HDR **à l'Université Paul Sabatier**

Monsieur Waseem SAFI présente, en vue de l'obtention du grade de docteur en Informatique de l'Université de Caen un mémoire intitulé : « Non-visual Vibro-tactile Navigation of Web Pages on Touche Screen devices ». Ce travail a été mené sous la direction de Messieurs Pierre BEUST et Jean-Marc ROUTOURE à l'Université de Caen avec un encadrement très important de la part de Fabrice MAUREL, notamment responsable du projet ART-ADN dans lequel la thèse a été financée.

Le travail présenté s'inscrit dans le cadre de l'« e-accessibilité » appliqué au problème de l'accessibilité par des personnes non voyantes à la structure globale des pages web. L'étude porte particulièrement sur la proposition d'une approche basée sur une modalité vibro-tactile dans le contexte d'appareils mobiles avec écrans tactiles. L'objectif est (1) de concevoir un système qui convertit les structures visuelles d'une page web à des vibrations tactiles (un système de substitution sensorielle de la vision tactile), (2) d'élaborer une plateforme expérimentale, et (3) d'évaluer expérimentalement différentes hypothèses. Le document est composé de 213 pages, de 156 références bibliographiques.

Le mémoire soumis par M. Waseem SAFI est rédigé en anglais, il est agréable à lire, le style est très clair, les illustrations sont nombreuses (x Tableaux, de x figures). Il est divisé en six chapitres intitulés respectivement :

1. Access to Spatial Information in Web Pages by Visually Impaired Persons,
2. Non-visual Access to Visual Information in Numerical Documents,
3. TactiNET Framework, Hardware Implementation,
4. Software Implementation,
5. Basic Parameters and Arguments of the Graphical Vibro-tactile language,
6. Recognizing the Structures of Web Pages Through Vibro-Tactile Feedbacks.

auxquels s'ajoutent une introduction générale et une conclusion/perspectives qui rappellent les principaux résultats. Le candidat complète son document par 4 annexes.

Les chapitres sont regroupés en 3 parties :

1. Targeted Users, State of the Art, regroupant les trois premiers,
2. System Description, Hardware and Software implementation, regroupant les deux suivants,
3. Recognizing the Structures of Web Pages Through Vibro-Tactile Feedbacks, constitué par le dernier chapitre 6.

Dans **la première partie**, le candidat dresse un état de l’art, bien structuré et richement documenté, relatif à l’accès aux informations spatiales par les personnes aveugles (***chapitre 1***) et à l’accès non visuel dans les documents numériques (***chapitre 2***). Le premier chapitre est consacré aux concepts de l’accessibilité du web, aux limites des technologies d’assistance, et aux concepts de la perception des dispositions spatiales et des principes des sensations vibro-tactiles. Le deuxième chapitre décrit les concepts de la substitution des percepts de la vision vers le tactile et présente trois catégories de systèmes qui s’appuie sur ces concepts. Une attention particulière a été portée sur le critère de la mobilité (contexte spécifique de la thèse), les problèmes de la « réduction » et de la transformation de la densité des informations visuelles, et sur l’approche théorique du couplage « Action-Perception » et l’ergonomie des interactions.

La **deuxième partie** est consacrée à une présentation très détaillée de la TactiNET (dans ses deux versions), un dispositif vibrotactile permettant de percevoir, sur tablette de type Androïd, les différents contrastes lumineux produits à l’écran. Le but de la plateforme de la TectiNET est d’étudier les effets des retours vibro-tactiles sur la capacité des personnes non voyantes à percevoir la mise en page d’une page web lors de la navigation sur un écran tactile.

Le ***premier chapitre*** présente les deux versions de l’implémentation matérielle. Les critères, et les conditions fixées pour l’implémentation correspondent à la portabilité, la réduction du coût, l'utilisabilité, la luminosité, le bruit ambiant, un support sans fil, une utilisation en temps réel, et l’indépendance au codage Braille. Une attention particulière a été porté à la conception d'une interface homme machine, en tenant compte des spécifications de l'équipe partenaire en psychologie et facilitant ainsi son utilisation par les expérimentateurs. Dans la deuxième version de la TactiNET, plus aboutie, il est possible de contrôler les paramètres de fréquence et d'amplitude des actionneurs. Ces fréquences et amplitudes représentent les différents segments et zones présents dans une page web.

Le ***deuxième chapitre*** présente les ressources logicielles pour obtenir cette extraction automatiquement. Le principe de regroupement et de clustering des segments d’une page HTML se base (1) sur la théorie de la Gestalt. L’algorithme proposé par la TactiNET se base sur les techniques de traitement d’image (morphologie) appliqué dans les cas des structures géométriques ; et (2) sur la possibilité de paramétrer le nombre de regroupements souhaité dans une page. L’algorithme est organisé en trois phases : l’extraction des paramètres HTML, le filtrage et l’organisation de ces paramètres, et enfin le regroupement. 154 pages web (sites d'information, de e-commerce et touristiques) ont été étudiées et observées « manuellement » avant d'implémenter cet algorithme. Un travail nécessaire et utile pour garantir un bon résultat. La première phase a été le résultat d’une collaboration avec l’entreprise Semio Time dans le cadre d’un projet ANR (Art-ADN).
Une remarque à propos de la deuxième phase est que certains paramètres ont été filtré alors qu'ils renseignent sur le visuel des éléments logique de la page.
Deux remarques à propos de la troisième phase : (1) il faudra garder une trace de la finesse de la segmentation pour une utilisation postérieure avec éventuellement une modalité orale ; (2) la proximité spatiale n’est pas la seule propriété à utiliser pour un regroupement (deux segments distants peuvent être associés).

Pour illustrer la difficulté de la segmentation et faire des comparaisons, une étude empirique a été menée auprès de 15 sujets sur un corpus de 480 pages. J'ai cependant un petit regret quant à l'absence d'argumentaire et d'une classification des erreurs dans cette phase de regroupement. Ces remarques ont pour objet essentiel d'inviter à une discussion pendant la soutenance avec le candidat et n'enlève rien à la qualité et à la quantité de travail accompli par le candidat.

Le choix des pages représentatives des trois catégories de sites (information, e-commerce et touristique) sur lesquelles ont porté toutes les expérimentations a été opéré avec méthodologie et l'approche de calcul de similarité de la mise en page qui a été choisie est celle de [Zhang et al., 2012] qui se base sur les caractéristiques des éléments HTML d'une page web qui sont vus comme des boîtes rectangulaires.
Le ***chapitre 5*** et dernier de la partie 2 présente les paramètres et les éléments du langage graphique vibr-tactile et argumentent les différents choix.
La méthodologie pour définir les protocoles, mener les expériences et évaluer les résultats ont été élaboré avec rigueur. Trois expériences ont été menées : d'une part sur la reconnaissance des formes visuelles par un retour vibo-tactile, et d'autre part sur la sélection de la fréquence et de l'amplitude les plus perceptibles et les plus discriminantes des contrastes visuels. Ces premières expériences ont été mené avec la première version de la TactiNET (un seul actionneur). Des analyses Anova sont réalisées pour étudier les effets significatifs des variables.
Des résultats importants ont été trouvés et résumés dans ce dernier chapitre notamment sur la fréquence optimale de perception et de discrimination et également sur des compartifs entre voyants et non voyants.
Une remarque concerne la durée qui est un peu longue sur certaines passations (jusqu'à deux heure et demie).

La **troisième partie** est constitué d'un seul chapitre qui présente une série d'expériences d'une grande importance dans cette thèse et qui ont pour objectif (1) d'étudier la capacité de discrimination des structures des pages web à partir des retours vibro-tactiles et (2) d'étudier l'image mentale que se sont construites les participants après la navigation. Ces expériences sont conduites auprès de 6 sujets non voyants entre 39 et 69 ans, moitié femme et moitié homme et dont la moitié ont perdu la vue à la naissance, et 5 sujets voyants (femmes) entre 25 et 29 ans.
D'abord, pour le choix des pages représentatives des trois catégories de sites (information, e-commerce et touristique) un algorithme a été appliqué sur 900 pages et les 3 patterns de vibrations ont été associés aux 5 zones semi-automatiquement.
Une remarque sur le fait de ne pas corriger manuellement toutes les erreurs du résultat de l'algorithme de segmentation et de regroupement avant d'effectuer l'expérimentation.
Une deuxième remarque sur l'adaptation des segmentations des pages pour l'expérimentation qui ne tient pas compte des dimensions de manière proportionnelle. C'est le cas par exemple de la page représentative de la catégorie touristique où la zone 2 devrait être plus grande que la zone 1.
A la suite des expériences de reconnaissances des structures, il était demandé aux participants de redessiner sur un papier deux des trois pages web montrées pendant l'expérience.
Et pour évaluer le pourcentage de précision des formes redessinées, elles ont été présentées à des experts.
Une section exhibe un résultat intéressant sur les stratégies de navigation (12). Ces stratégies concernent à la fois des navigations en locale (NL) et en globale (NG). Un travail

complémentaire pourrait venir affiner ce résultat pour distinguer lesquelles concernent exclusivement une NL ou exclusivement une NG ou les deux, et dans quels patterns de situation.
L'auteur présente dans la conclusion 15 résultats très intéressants. Pour citer quelques exemples : (1) 62.1% du total des participants reconnaissent les structures des pages web ; (2) plus les participants sont familiers avec les techniques des écrans tactiles plus ils sont capables d'utiliser le système de la TactiNET ; (3) il y une corrélation significative pour la majorité des participants entre leur performance dans la tâche de reconnaissance et la tâche de reproduction de l'image mentale ; (4) la majorité des participants utilisent au moins deux stratégies de navigation parmi les 12 répertoriées ; (5) deux stratégies particulières en navigation locale, représentées par des stratégies verticale et horizontale continues, sont les stratégies les plus utilisées par la majorité des participants. Il s'agit de parcours qui s'apparentent à une écriture en mode Boustrophédon.
Un petit regret sur l'absence de phase d'apprentissage qui n'a pu avoir lieu pour des raisons de temps et de délai de remise de livrable.
L'auteur termine ce dernier chapitre par ceci « Analyzing many parameters concerning the global and the micro navigation strategies such the time, the speed, and their successive might result new observations. These observations could be useful in developing new versions of the designed system.», ce qui ouvre de grandes perspectives à ce travail de qualité.

Le candidat clôture son mémoire par une conclusion générale et quelques perspectives. On ne peut que regretter le petit espace qui a été alloué à cette section qui mériterait plus d'argumentaires et de discussions.

En conclusion, il s'agit d'un travail conséquent à la fois expérimental et de modélisation, sur une problématique complexe necessitant des compétences pluridisciplinaires (informatique, électronique, interaction homme machine, psychologie cognitive) ayant abouti à des résultats intéressants. Le travail important d'expérimentation, la bonne connaissance des travaux du domaine, le choix des hypothèses de travail et des méthodes utilisées, permettent d'affirmer que M. Waseem SAFI présente une bonne maîtrise du sujet et apporte une contribution dans le domaine. De plus, le candidat a commencé une diffusion de ces travaux matérialisée par plusieurs publications d'audience internationale. En conséquence, je donne un avis favorable à la soutenance orale de ce travail en vue de l'obtention du titre de docteur de l'Université de Toulouse, dans la spécialité Informatique.

Toulouse, le 14 Novembre 2016

Mustapha Mojahid
Maître de conférences – HDR

Rapport sur la soutenance du diplôme de doctorat de l'université de Caen Normandie de Waseem SAFI

Caen, le 14/12/16

Spécialité : Informatique et applications
Prénom Nom : Waseem SAFI
Date de soutenance : mardi 14 décembre 2016
Président du Jury : Franck POIRIER, Professeur, Université de Bretagne Sud
Constitution du Jury : Pierre BEUST, Fabrice MAUREL, Mustapha MOJAHID, Michèle MOLINA, Thomas PIETRZAK, Jean-Marc ROUTOURE, Khaldoun ZREIK

Waseem Safi a bien montré la large couverture, l'originalité et la pluridisciplinarité du sujet. Son exposé, en français, a été bien structuré et bien illustré. A l'issue de l'exposé, les questions couvrant tous les champs disciplinaires de la thèse ont été très nombreuses et ont abordé différents volets du travail. Il a su répondre de manière réactive, satisfaisante et précise montrant une bonne maîtrise du sujet. Le jury a apprécié l'ampleur du travail de recherche et de développement fourni dans cette thèse ainsi que les efforts méthodologiques pour les expérimentations. Le jury souligne que son travail ouvre de réelles perspectives de recherche.

Pour toutes ces raisons, le jury à l'unanimité, décerne à Monsieur **Waseem Safi**, le titre de **docteur en informatique de l'université de Caen Normandie**.

Michèle Molina   Franck Poirier   Thomas Pietrzak

Jean-Marc Routoure   K. ZREIK   Mustapha MOJAHID

A Hybrid Segmentation of Web Pages for Vibro-Tactile Access on Touch-Screen Devices

Waseem Safi[1] **Fabrice Maurel**[1] **Jean-Marc Routoure**[1, 2] **Pierre Beust**[1] **Gaël Dias**[1]

[1] University of Caen Basse-Normandie - UNICAEN

[2] National Superior Engineering School of Caen - ENSICAEN

14032 Caen - France

`firstName.lastName@unicaen.fr`

Abstract

Navigating the Web is one of important missions in the field of computer accessibility. Many specialized techniques for Visually Impaired People (VIP) succeed to extract the visual and textual information displayed on digital screens and transform it in a linear way: either through a written format on special Braille devices or a vocal output using text-to-speech synthesizers. However, many researches confirm that the perception of the layout of web pages enhances web navigation and memorization. But, most existing screen readers still fail to transform the 2-dimension structures of web pages into higher orders. In this paper, we propose a new framework to enhance VIP web accessibility by affording a "first glance" web page overview, and by suggesting a hybrid segmentation algorithm to afford nested and easy navigation of web pages. In particular, the web page layout is transformed into a coarse grain structure, which is then converted into vibrating pages using a graphical vibro-tactile language. First experiments with blind users show interesting issues on touch-screen devices.

1 Introduction

In October 2013, the world health organization estimated that the number of Visually Impaired People (VIP) in the world is 285 million: 39 million of them are blind and 246 million have low vision. In particular, the organization defined four levels of visual functions depending on the international classification of diseases: normal vision, moderate visual impairment, severe visual impairment and blindness.

VIP depend on screen readers in order to deal with computer operating systems and computational programs. One of most important and desired targets by VIP is navigating the Web, considering the increased importance and expansion of web-based computational programs. Screen readers present some solutions to navigate the textual and graphical contents of web pages, either by transforming a web page into a written Braille, or into a vocal output. In addition to these solutions, some screen readers installed on touch devices transform a web page into a vocal-tactile output.

But there are some drawbacks for these proposed solutions. On the one hand, Braille techniques are costly and only a few number of VIP have learned Braille (in France in 2011, there were about 77,000 visually impaired people and only 15,000 of them had learned Braille). On the other hand, transforming the information of a web page into a vocal format might not be suitable in public and noisy environments. Finally most of Braille solutions are not suitable for mobile devices [1].

In addition to these drawbacks, the most important one is the failure to transform the 2-dimensions web page structure. Indeed, as reported by many authors, perceiving the 2D structure of web documents greatly improves navigation efficiency and memorization as it allows high level text reading strategies such as: rapid or cursory reading, finding or locating information, to name but a few [2].

Our work focuses on developing and evaluating a sensory substitution system based on a vibro-tactile solution, which may solve the mentioned drawbacks. In particular, we study how to increase the VIP perception of a 2D web page structure and how to enhance their techniques to navigate the contents of

web pages on touch-screen devices. The suggested solution is very cheap compared to Braille devices and may be efficient in noisy and public environments compared to vocal-tactile solutions.
Our contribution is three-fold: (1) designing a Tactile Vision Sensory System (TVSS) represented by an electronic circuit and an Android program in order to transform light contrasts of touch-screen devices into low-frequencies tactile vibrations; (2) designing an algorithm for segmenting web pages in order to support the visually impaired persons by a way which may enhance their ability to navigate the textual and graphical contents of web pages and (3) analyzing the effects of the suggested segmentation method on navigation models and tactics of blind persons, and its effect on enhancing their strategies for text reading and looking for textual information.

The paper is organized as follows. First, in section 2, we review most advanced VIP targeted technologies. Then, in section 3, we describe the new proposed framework. In section 4, we view the state of the art for web pages segmentation methods. In the fifth section, our hybrid segmentation method is presented and how this method could be integrated in our framework. In section 6, we enumerate the desired effects of the proposed segmentation method on navigation models and tactics of blind persons, and how it may enhance their strategies for text reading and searching of textual information. Finally, in the seventh section, perspectives and conclusions are presented.

2 VIP targeted technologies

Current products for VIP such as screen readers mainly depend on speech synthesis or Braille solutions, e.g. ChromeVox [3], Windows-Eyes [4], or Job Access With Speech [5]. Braille displays are complex and expensive electromechanical devices that connect to a computer and display Braille characters. Speech synthesis engines convert texts into artificial speech, where the text is analyzed and transformed into phonemes. These phonemes are then processed using signal processing techniques.

Some screen readers can also support tactile feedback when working on touch-screen devices, such as Mobile Accessibility [6] and Talkback [7] for Android, or VoiceOver [8] for IPad. Many of these products propose shortcuts for blind users to display a menu of HTML elements existing in the Web page, for example headers, links and images. But, the main drawback of all these products is the fact that they transfer the information of web pages into a linear way i.e. without any indication of the global structure.
Many researches tried to enhance the way by which VIP interact with web pages, such as [9], who proposed a tactile web navigator to enable blind people to access the Internet. This navigator extracts texts from web pages and sends them to a microcontroller responsible for displaying the text in Braille language using an array of solenoids. A tactile web browser for hypertext documents has been proposed by [10]. This browser renders texts and graphics for VIP on a tactile graphics display and supports also a voice output to read textual paragraphs and to provide a vocal feedback. The authors implemented two exploration modes, one for bitmap graphics and another one for scalable vector graphics. A pin matrix device is then used to produce the output signal for blind users. The main drawback of these two proposed systems is that they need specific devices (solenoids and pin matrix), which are expensive and cannot be integrated to handled devices such as PDAs or Tablet PCs. Another interesting model called MAP-RDF (Model of Architecture of Web Pages) is proposed by [11]. This model allows representing the structure of a web page and provides blind users with an overview of the web page layout and the document structure semantics. Tactos is a perceptual interaction system, which has been suggested by [12] and consists of three elements: (1) tactile simulators (two Braille cells with 8 pins) represent a tactile feedback system, (2) a graphics tablet with a stylus represents an input device and (3) the computer. More than 30 prototypes of Tactos have been released to serve a lot of users in many domains. Tactos has been successfully used to recognize simple and complex shapes. The device has been also used in geometry teaching domain in an institution for visually impaired and blind children. Tactos also allowed psychology researchers to propose and develop new paradigms for studying perceptions and mediated communication of blind persons [13]. However, it shows the same drawback as the previous systems, which are expensive and need specific devices. Moreover, the blind user can only explore the web page with a stylus and both hands are occupied by the system. Moreover, it is unemployable for a large set of environments, for example in public.

3 Proposed Framework

The "first glance" can be defined as the ability to understand the document layout and its structural semantics in a blink of an eye [1]. In this work, we aim to increase the ability of visually impaired persons to understand the web page 2-dimensions layout in order to enhance their tactics to navigate the Web with a vibro-tactile feedback.

The first phase in our model is to extract visual structures in the navigated web page and convert these "visual" blocks into zones (or segments) to facilitate the navigation in later phases. We achieve this phase depending on a hybrid segmentation method. Then the system represents the extracted visual elements as symbols using a graphical language. The third phase is to browse these graphical symbols depending on the size of the used touched-screen device; and in the fourth phase, our system provides a vibro-tactile feedback when the blind user touches the tablet by giving the user a vibro-tactile feedback by transforming light contrasts of touch-screen devices into low-frequencies tactile vibrations. A tablet (Asus Model TF101 with Android operating system) has being used for our tests.

To achieve the desired system, we have designed an electronic circuit, which controls two micro-vibrators placed on two fingers. A Bluetooth connection with an android tablet allows controlling the vibration intensity (i.e. amplitude) of vibrators. An Android dedicated program on the tablet views an image on the screen and detects where the user touches the tablet screen (the viewed image represents the result of web page segmentation). The intensity of the light emitted by the tablet at touched points is then transmitted to the embedded device in order to control the vibration intensity. In this paper, we focus only on the first phase, with considering that the detailed description of the system, and results of pre-tests are described in [1] and [14].

4 Related Works

Segmenting a web page is a fundamental phase for understanding its global structure. Extracting the global structure of web pages is useful in many domains such as information retrieval, data extraction, and similarity of web pages.

Many approaches have been suggested for segmenting web pages, such as:
-1) DOM-based segmentation: it depends on analyzing the DOM tree (Document Object Model), and extracting the main structure of web pages depending on HTML tags. An example of this approach is the work of [15], which determines firstly the layout template of a web page, and then it divides the page into minimum blocks, and finally collects these minimum blocks into content blocks.
-2) Vision-based segmentation: this method divides the web page depending on the visual view of web page contents on a web browser. The most famous tool depends on this approach is VIPS (VIsion based Page Segmentation) [16]
-3) Image processing based segmentation: this approach depends on capturing an image for the visual view of a web page, and then depends on image processing techniques to divide the captured image into sub blocks [17] [18].
-4) Text-based Segmentation: this approach focuses on extracting only information about texts existed in a web page. After dividing the web page into blocks of texts, it could be possible to find the semantic relations between these textual blocks. This method is very useful in many information retrieval domains such as question answering applications [19].
-5) Fixed-length segmentation: this approach divides the web pages into fixed length blocks (passages), after removing all HTML tags, where each passage contains a fixed number of words [20].
-6) Densitometric analysis based segmentation: this approach depends on methods applied in quantitative linguistics, where text-density refers to a measure for identifying important textual segments of a web page [21].
-7) Graph-based segmentation: This approach depends on transforming the visual segments of a web page into graph nodes, then applying many common graph methods on these nodes for combining them into blocks, or for making a clustering for these nodes. Some common works which depend on this approach are [22] [23].
-8) and Hybrid-based segmentation: This approach combines many approaches indicated previously.

5 Suggested Hybrid Segmentation Algorithm

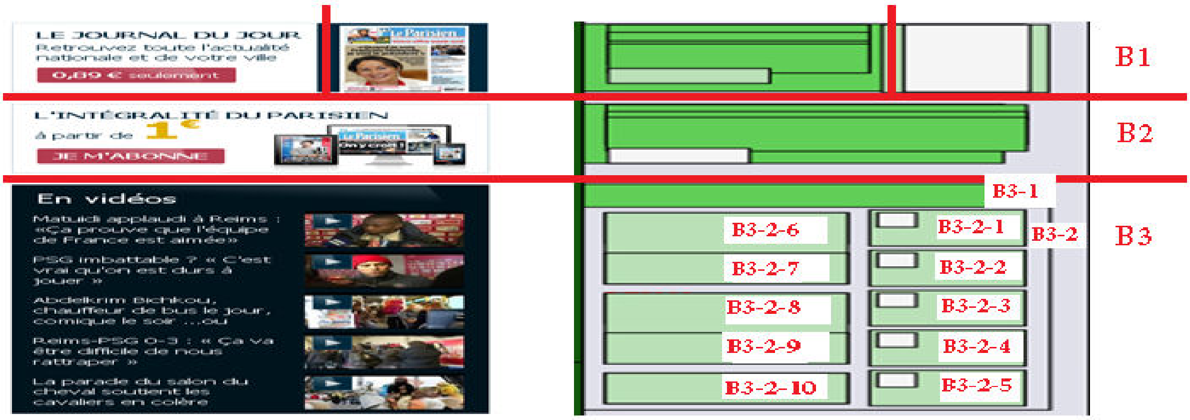
Most of segmentation algorithms render firstly the web page using a web browser, and then segments the HTML elements into many blocks depending on the visual layout. The constructed hybrid segmentation algorithm has been tested on 154 pages collected manually from many information and e-commerce sites (www.leparisien.fr, www.lefigaro.fr, www.liberation.fr, www.amazon.fr, www.materiel.net), and we have integrated the results with our under-development Android program. The obtained results are promised because the segmentation algorithm can extract well the web page blocks depending on the visual structure, and the algorithm can also convert correctly these blocks into zones (clustering the blocks). Our algorithm blends three segmentation approaches, DOM-based segmentation, vision-based segmentation, and graph-based segmentation.

5.1 Proposed Corpora

To achieve the previous mentioned model, we are constructing two corpora, one for training, and another for testing. We selected many criteria for crawling web pages, such as, the type of crawled pages (information web sites, and e-commerce web sites), the size (about 10,000 pages), the language (French), the version of web site (Classic, Mobile), and the technology used to build the crawled web site (framework JavaScript: JQuery, mootools, ... / CMS: Prestashop, Drupal, Joomla...).

5.2 Vision-Based Approach

In this phase, we render the web page using Mozilla FireFox browser, and getting its visual structure by injection Java-script code inside the HTML source of the rendered web page. The visual structure indicates a global hierarchy of the rendered web page, and indicates the bounding box for each HTML element. Figure 1(a) represents a part of a web page, and the result of its vision-based segmentation is illustrated in figure 1(b).



(a) A part of a web page (b) Vision-based segmentation

Figure 1. A part of a web page (leparisien.fr) and its vision-based segmentation

The input of this phase is a web page HTML source, and its output is injected information about bounding boxes for each HTML element. In next sections, we refer to bounding boxes by blocks (i.e. each bounding box represents an HTML element, and may contain many bounding boxes.).

5.3 DOM-Based Approach

After segmenting a web page depending on its visual structure, we analyze its DOM structure by applying filters and re-organization rules for enhancing results of next phases. Dead-Nodes filter is an example of these filters: it deletes all HTML nodes that do not affect on the appearance, for example nodes with height or width equals to “0px” (zero pixel); or nodes with style properties ("display : none" or "visibility:hidden"). An example of re-organization rules is Paragraph-Reorganization rule, where this rule re-constructs all paragraph child-nodes in one node contains the extracted text; we made this rule after analyzing many DOM structures, and observing that some paragraph nodes contain child-nodes which affect negatively on extracting the text, such as <i>, <span>, etc.., and these child-nodes contain important texts. We had put many filters and re-organization rules, and integrated them with our framework, and then we tested applying these rules and filters on the vision-based segmented web pages (154 pages mentioned previously). As a result of applying the two approaches (vision-based and DOM-based), we succeeded to get the first glance visual structure for many pages.

The Result of this phase is a filtered DOM-tree, each of its nodes is visible and contains a bounding box information. Figure 1 represents a hierarchy of some HTML nodes, there are 3 main blocks (B1, B2, and B3), and each one contains many sub blocks, for example B3 contains B3-1 and B3-2.

To illustrate results of applying previous two mentioned approaches, we represented a resulted filtered DOM-tree on the used tablet. Figure 2 views a graphical representation for a page web, since each rectangle represents a block in the analyzed web page. Red rectangles represent images (<img> tags), green rectangles represent links (<a>), blue rectangles represent list of items (<il> or <ol> tags), black rectangles represent paragraphs (<p> tags).

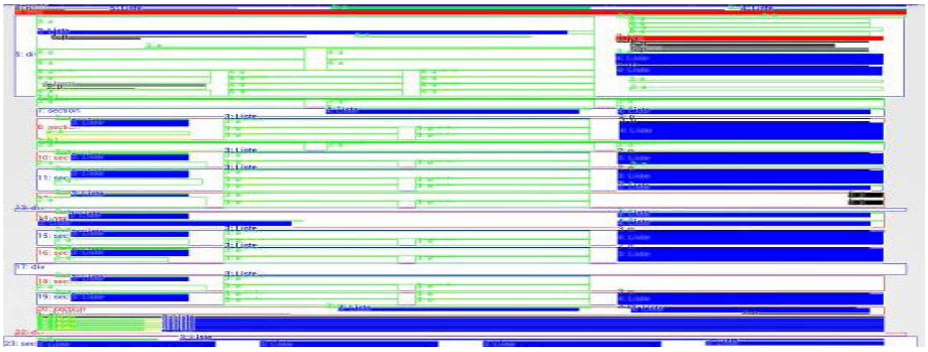


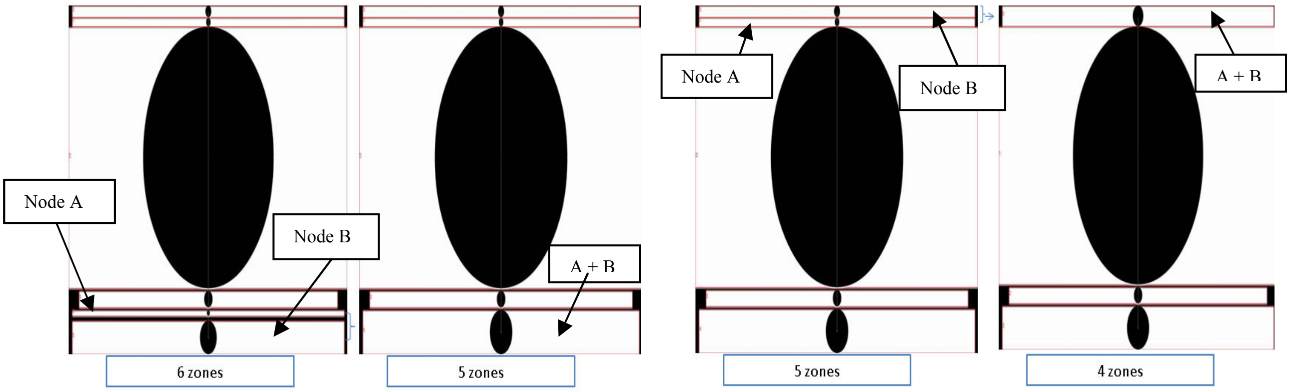
Figure 2. A graphical representation for a page web (leparisien.fr)

5.4 Graph-Based Approach

After segmenting the web page depending on its visual structures and analyzing its DOM-structure, we apply a new graph-based segmentation algorithm which called "Blocks2Zones Clustering" in order to group many blocks together in one zone. Clustering many blocks together is necessary to decrease the number of viewed blocks in some interfaces (instead of viewing many blocks, we view one zone represents these blocks and then the user can navigate intra-elements inside the zone by double clicking on the graphical element of the zone.), and to group closed blocks in one zone (here, closeness depends on distances between blocks, this will be described next sections in details). The pseudo-code of the proposed algorithm is:

Blocks2Zones Clustering Algorithm
Input (Blocks, N° of desired Zones)
Output: Graph of N nodes (N Zones)
1- Transform the blocks into a graph (Non-Directed graph)
1.1. Blocks → Nodes,
1.2. Make relations between the nodes, and assign weights for these relations.
2- If number of zones >= number of blocks
end the algorithm,
Else
3- Find the node with the smallest size (node A) (Figure 3.a (6 zones), Figure 3.b (5 zones))
4- For the node A, find the relation which has the largest weight (node B). (Figure 3.a (6 zones), Figure 3.b (5 zones))
5- Group the nodes A, and B. (Figure 3.a (5 zones), Figure 3.b (4 zones))
6- Repeat the steps 3-4-5 till number of blocks == number of zones

Figure 3 represents some examples of applying this algorithm, where each rectangle represents a Zone (a block or collection of blocks), and the center of each ellipse represents the zone center.



(a) Converting 6 zones to 5 zones (b) Converting 5 zones to 4 zones

Figure 3. Examples of applying Blocks2Zones clustering Algorithm

To calculate weights between nodes, we tested 2 relations of distances: the first one is Minkowski Manhattan distance ($d(p,q) = d(q,p) = \|p - q\| = \sum_{i=1}^{n} |p_i - q_i|$), and the second is Minkowski Euclidian distance ($d(p,q) = d(q,p) = \sqrt{(p_1 - q_{1)}{}^2 + (p_2 - q_{2)}{}^2 + \ldots + (p_n - q_{n)}{}^2}$). To ensure which distance should be used, we applied an internal quality criterion for the two used distances; the applied criterion is Sum of Squared Error (SSE) ($SSE = \sum_{K=1}^{n} \sum_{\forall x_i \in C_K}^{n} ||x_i - \alpha_i||^2$ Where Ck is the set of instances in cluster K, and $\alpha_{k,j} = \frac{1}{N_k} \sum_{\forall x_i \in C_K}^{n} x_{i,j}$).

Results of applying SSE measure for the two distances (Minkowski Manhattan, and Minkowski Euclidian) were the same; which means that using either Minkowski Manhattan distance or Minkowski Euclidian distance is equal in our algorithm to calculate weights between nodes.

Applying this hybrid segmentation algorithm on a web page converts it to a set of zones, each zone contains many other zones or blocks, and each block represents a visual structure and may contain many other blocks. The purpose of the proposed vibro-tactile access protocol is then to transform the semantic of symbols in these zones, or blocks, or HTML elements into vibrations of different frequencies and amplitudes.

6 Desired Effects of Suggested Algorithm on Web Navigation Models of VIP

We had finished designing the suggested algorithm, and integrated it with our framework. However, practical experiments with VIP are the most important criteria to select success or failing this algorithm in enhancing web navigation models of VIP, and this is our next step to be achieved. But, depending on our previous experiments in dealing with VIP targeted techniques, we can expect some desired effects of the proposed segmentation method on navigation models and tactics of blind persons, and how it may enhance their strategies for text reading and searching of textual information.

Firstly, this segmentation model can give VIP an impression of the layout of navigated web page (first glance structure), and (as indicated previously) perceiving the 2D structure of web documents greatly improves navigation efficiency and memorization.

Secondly, this suggested segmentation method can group together many closed blocks in one zone, and in this way the user can select easily if this zone contents are important for him/her or not, for example collecting all header blocks in one zone, or collecting all footer blocks in one zone.

Thirdly, this model can allow high level text reading strategies such as rapid or cursory reading or locating information, this can be achieved by ignoring navigating any zone does not contain textual information. By the way if one zone contains textual information, the user can navigate it and decide if it contains important information for him/her.

7 CONCLUSION AND PERSPECTIVES

In this paper, we summarized our current work which aims to design an approach for non-visual access to web pages on touch-screen devices, and focused on the suggested hybrid segmentation algorithm. We expect that integrating this method of segmentation with the designed vibro-tactile protocol can give VIP an impression of the first glance layout of web pages.

In the same way that the environment enables a blind person to move in space with sidewalks and textures which will be explored by his/her white cane, we hope giving the blind user an ability to navigate documents depending on "textual sidewalks" and "graphical paths" which will be discovered by his/her finger.

Next steps in this research will be 1) making real experiments to study effects of suggested segmentation algorithm on web navigation models of VIP, 2) adding advances techniques in text summarization to facilitate navigating textual information, 3) Adding elements to the graphical vibro-tactile language in order to represent more HTML elements such links, buttons, input fields, and other elements, 4) We plan also to add thermic actuators for translating the notion of colors. This may be very useful and hopeful for blind users to transfer information about colors.

Acknowledgment

This works is founded by the national agency of research ANR (Agence Nationale de Recherche http://www.agence-nationale-recherche.fr/) as a part of ART-ADN project ANR-12-SOIN-0003-02 (Accès par Retour Tactilo-oral Aux Documents Numériques) in GREYC laboratory (www.greyc.fr) of the University of Caen Basse-Normandie (www.unicaen.fr) and the National Superior Engineering School of Caen (www.ensicaen.fr).

8 References

[1] Maurel, F., Dias, G., Routoure, J-M., Vautier, M., Beust, P., Molina, M., Sann, C., 2012. « *Haptic Perception of Document Structure for Visually Impaired People on Handled Devices* », Procedia Computer Science, Volume 14, 2012, Pages 319-329, ISSN : 1877-0509. DOI=http://dx.doi.org/10.1016/j.procs.2012.10.036

[2] Maurel, F., Vigouroux, N., Raynal, M., Oriola, B., 2003. « *Contribution of the Transmodality Concept to Improve Web Accessibility* ». In Assistive Technology Research Series, Volume 12, 2003, Pages 186-193. International conference; 1st, Smart homes and health telematics; Independent living for persons with disabilities and elderly people. ISSN : 1383-813X.

[3] http://www.chromevox.com/ [Access 24/5/2014]

[4] http://www.synapseadaptive.com/gw/wineyes.htm [Access 24/5/2014]

[5] http://www.freedomscientific.com/ [Access 24/5/2014]

[6]https://play.google.com/store/apps/details?id=es.codefactory.android.app.ma.vocalizerfrfdemo&hl=fr [Access 24/5/2014]

[7]https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=fr [Access 24/5/2014]

[8] http://www.apple.com/fr/accessibility/ [Access 24/5/2014]

[9] Alaeldin, A., Mustafa , Y., Sharief, B., 2012. « *Tactile WebNavigator Device for Blind and Visually Impaired People* ». In Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies, Jordan, 2012.
DOI=http://dx.doi.org/10.1109/AEECT.2011.6132519

[10] Rotard, M., Knödler, S., Ertl, T., 2005. « *A Tactile Web Browser for the Visually Disabled* ». In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, 2005, pages 15-22.
DOI= http://dx.doi.org/10.1145/1083356.1083361

[11] BouIssa , Y., Mojahid, M., Oriola, B., Vigouroux, N., 2009. « *Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents* ». IEEE International Conference on Advances in Computational Tools for Engineering Applications, 2009, Pages: 591 – 594.
DOI=http://dx.doi.org/10.1109/ACTEA.2009.5227855

[12] Lenay, C., Gapenne, O., Hanneton, S., Marque, C., Genouëlle, C., "*Sensory Substitution, Limits and Perspectives*». In Touch for Knowing Cognitive psychology of haptic manual perception, Amsterdam, 2003, Pages: 275-292.

[13] Tixier, M., Lenay, C., Le-Bihan, G., Gapenne, O., Aubert, D., « *Designing Interactive Content with Blind Users for a Perceptual Supplementation System* », TEI 2013, 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain, 2013, Pages 229-236.
DOI= http://dx.doi.org/10.1145/2460625.2460663

[14] MAUREL, F., SAFI, W., BEUST, P., ROUTOURE, J.M., 2013. « *Navigation aveugle sur dispositifs mobiles : toucher le Web... pour mieux l'entendre*», 16ème Colloque International sur le Document Électronique, CIDE16, Lille France, Europia productions, 2013.

[15] Sanoja, A., Gançarski, S., «*Block-o-Matic: a Web Page Segmentation Tool*», BDA. Nantes, France. 2013.
http://hal.archives-ouvertes.fr/hal-00881693/

[16] Deng, C., Shipeng, Y., Ji-Rong, W., Wei-Ying, M., «*VIPS: a Vision-based Page Segmentation Algorithm*», Nov. 1, 2003, Technical Report MSR-TR-2003-79, Microsoft Research. 2003.
http://research.microsoft.com/pubs/70027/tr-2003-79.pdf

[17] Cai, D., He, X., Ma, W-Y., Wen, J-R., Zhang, H., 2004. « *Organizing WWW Images Based on the Analysis Of Page Layout And Web Link Structure*», Microsoft Research Asia, Beijing, China, 2004.
http://research.microsoft.com/pubs/69080/25.pdf

[18] Cao, J., Mao, B., Luo, J., 2010. « *A segmentation method for web page analysis using shrinking and dividing*», International Journal of Parallel, Emergent and Distributed Systems - Network and parallel computing, Volume 25 Issue 2, April 2010. Pages: 93-104.
DOI=http://dx.doi.org/10.1080/17445760802429585

[19] Foucault, N., Rosset, S., Adda, G., 2013. « *Pré-segmentation de pages web et sélection de documents pertinent en Questions-Réponses*», TALN-RÉCITAL 2013.

[20] Callan, J.P., 1994. « *Passage- level Evidence in Document Retrieval*», the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, 1994. Pages: 302-310. Publisher: Springer-Verlag New York, Inc.
http://dl.acm.org/citation.cfm?id=188589

[21] Kohlschütter, C., Nejdl, W., « *A Densitometric Approach to Web Page Segmentation*», USA, CIKM'08, Proceedings of the 17th ACM conference on Information and knowledge management, 2008. Pages: 1173-1182.
DOI: http://dx.doi.org/10.1145/1458082.1458237

[22] Chakrabarti, D., Kumar, R., Punera, K., 2008. «*A graph-theoretic approach to webpage segmentation*». Proceedings of the 17th international conference on World Wide Web, WWW'08, ACM, USA, 2008. Pages: 377-386. Publisher: ACM New York, NY, USA ©2008
DOI: http://dx.doi.org/10.1145/1367497.1367549

[23] Liu, X., Lin, H., Tian, Y., 2011. "*Segmenting Webpage with Gomory-Hu Tree Based Clustering*", Journal of Software, Vol 6, No 12, Pages: 2421-2425. 2011.

Blind Browsing on Hand-Held Devices: Touching the Web... to Understand it Better

Waseem Safi[1] Fabrice Maurel[1] Jean-Marc Routoure[1, 2] Pierre Beust[1] Gaël Dias[1]

[1] University of Caen Basse-Normandie – UNICAEN
[2] National Superior Engineering School of Caen - ENSICAEN
14032 Caen- France - +33(0)231567336 +33 (0)231452722

{waseem.safi, fabrice.maurel, jean-marc.routoure, pierre.beust, gael.dias}@unicaen.fr

ABSTRACT

Navigating the Web is one of important missions in the field of computer accessibility. Many specialized techniques for VIP (Visually Impaired People) succeeded to extract the information displayed on digital screens and succeeded to transform this information in a linear way either into a written format on special Braille devices, or into a vocal output using text to speech synthesizers. However, although this success, screen readers failed to transform the 2-dimentional structure of the navigated web page; despite many researches confirm that perception the structure enhances web navigation and memorization. In this paper, we propose a new technique aimed to enhance the VIP ability to navigate the Web by affording a “first glance” web page overview. This technique focuses on improving non-visual vibro-tactile access to web pages on touch-screen devices, based on extraction and re-organization the structure of texts and graphical elements for web pages, reformatting and converting automatically these visual structures and textual information into vibrating pages using a graphical vibro-tactile language.

Categories and Subject Descriptors

[H.5.2] [**User Interfaces**] Graphical user interfaces (GUI), Haptic I/O, User-centered design, User interface management systems.

General Terms

Algorithms, Design, Human Factors, Languages.

Keywords

Visually impaired people, non-visual access, graphical vibro-tactile language, visual structures, textual information.

1. INTRODUCTION

In October 2013, the world health organization estimated that the number of VIP in the world is 285 million, 39 million of them are blind, and 246 million of them have low vision. The organization defined four levels of visual functions depending on the international classification of diseases, which are: normal vision, moderate visual impairment, severe visual impairment, and blindness*.

VIP depend on screen readers in order to deal with computer operating systems and computational programs. One of most important and desired targets by VIP is navigating the Web, considering the increased importance and expansion of web-based computational programs. Screen readers present some solutions to navigate the Web, either by transforming a web page into a written Braille, or into a vocal output. Some screen readers installed on touch devices transform a web page into a vocal-tactile output. But there are some drawbacks for these proposed solutions: on the one hand, the Braille techniques are costly, and only few number of VIP have learned Braille (in France, there are about 77 000 visually impaired people and only 15 000 of them have learned Braille -from statistics published in September 2011-) **. On the other hand, transforming the information of a web page into a vocal format might not be suitable in public and noisy environments. Finally most of Braille solutions are not suitable for mobile devices [1]. In addition to these drawbacks, the most important one is the failure to transform the 2-D web page structure, because as reported by many authors, perception the 2D structure greatly improves navigation efficiency and memorizing the information because it allows high level reading strategies (rapid or cursory reading, finding or locating information,...) [2]. Our work focuses on developing and evaluating a sensory substitution system based on vibro-tactile solution which may solve the mentioned drawbacks; where we study how to increase the VIP perception of a 2-D web page structure, and how to enhance their techniques to navigate the Web on touch-screen devices. This suggested solution is very cheap comparing with prices of Braille devices, and also it could be more efficient in noisy and public environments comparing with vocal-tactile solutions. Our contribution is three-fold:

- Designing a Tactile Vision Sensory System (TVSS) represented by an electronic circuit and an android program in order to transform light contrasts of touch-screen devices into low-frequencies tactile vibrations,

- Running a series of experiments with blind persons in order to validate our hypotheses, and

- Analyzing many navigation models and tactics of blind persons.

The paper is organized as following: firstly, in section 2 we view the state of the art for VIP targeted technologies, and then we describe theoretical and methodological approaches for the new proposed technique, these approaches will be presented in section 3. The first pre-tests achieved with blind persons will then be described in section number 4, and analysis of results will be presented in the fifth section. Finally, in the last section the conclusion and perspectives will be proposed.

* http://www.who.int **http://www.opc.asso.fr/

2. STATE OF THE ART FOR VIP TECHNOLOGIES

Current products for VIP such as screen readers depend mainly on speech synthesis or Braille solutions, such as ChromeVox [3], Windows-Eyes [4], and Jaws (Job Access With Speech) [5]. Braille displays are complex and expensive electromechanical devices that connect to a computer and display Braille characters. Speech synthesis engines convert texts into artificial speech, where the text is analyzed and transformed into phonemes; these phonemes are then processed using signal processing techniques.

Some screen readers can support a tactile feedback when working on touch devices, such as Mobile Accessibility [6], Talkback [7] for Android, and VoiceOver [8] for IPad. Many of these products propose shortcuts for the blind user to display a menu of HTML elements existed in the web page, for example headers, links, and images. But, the main drawback of all these products is that they transfer the web page information into a linear way, and without any indication for the global web page structure (2D layouts).

Many researches tried to enhance the way by which VIP interact with web pages, such as [9], that proposed a tactile web navigator to enable blind people to access the Internet. This navigator extracts texts from web pages, and sends these texts to a microcontroller responsible of displaying the text in Braille language using an array of solenoids.

A tactile web browser for hypertext documents has been proposed by [14]. This browser renders texts and graphics for VIP on a tactile graphics display, and it supports a voice output to read textual paragraphs and to provide a vocal feedback. The authors implemented two exploration modes, one for bitmap graphics, and another one for Scalable Vector Graphics. Main drawback of this proposed system is that it needs a pin matrix device, which is expensive and cannot be integrated with handled devices.

Another interesting model called MAP-RDF ("Model of Architecture of web Pages") [10] proposed a method to improve the accessibility to visual information for blind persons. This model allows representing the structure of a web page, and provides the blind users with an overview of the web page layout and the document structure semantics. The main drawback of this model is that it could be applied only on well structured web pages which contain meta-data, so it could not be applied to most web pages which rarely contain meta-data. This model transforms the HTML elements to graphical symbols as illustrated in figure 1.

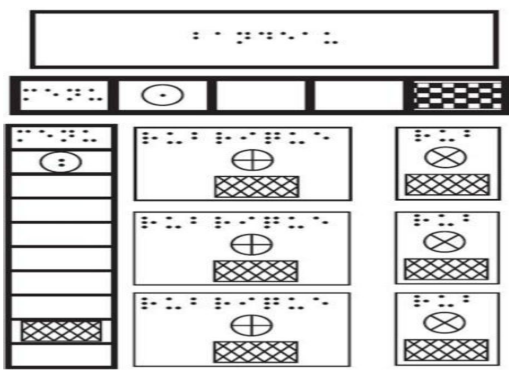


Figure 1. Representation of a web page by MAP-RDF model, figure extracted from [10].

In figure1, we notice many symbols; each one represents an HTML element. For example, the symbol represents a menu of items. And the symbols and represent texts with cold and hot colors.

Tactos is a perceptual interaction system suggested by [11], it consists of three elements: 1- tactile simulators (two Braille cells with 8 pins) represent a tactile feedback system, 2- a graphics tablet with a stylus (represents an input device), 3- computer [12], as shown in figure 2.

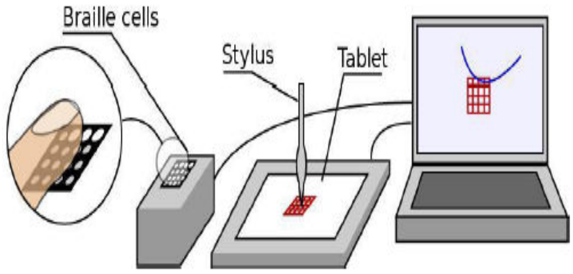


Figure 2. Different devices used in Tactos (Computer, graphics tablet, tactile simulators), figure extracted from [12].

The graphics tablet and the stylus allow the user to explore graphical contents on the screen such as circles, rectangles, and characters. While the user explores the contents, the system transforms pixels under the stylus into tactile stimulation on the Braille cells.

30 prototypes of Tactos have been released, to be used by a lot of users in many domains. Tactos has been successfully used to recognize simple and complex shapes. The device has been also used in geometry teaching domain in an institution for visually impaired and blind children. Tactos also allowed psychology researchers to propose and develop new paradigms for studying perceptions and mediated communication of blind persons [12].

3. PROPOSED TECHNIQUE

First glance could be defined as the ability -in a blink of an eye- to understand the document layout and its structural semantics [1]. We aim of our work to increase the ability of visually impaired persons to understand the web page 2-dimentional layout in order to enhance their tactics to navigate the Web. A commercial tablet connected to a vibro tactile set-up is used for that.

The first phase in our model is to extract visual structures in the navigated web page, and to convert these visual blocks to zones (segments) for facilitating the navigation in later phases. We achieve this phase depending on a hybrid segmentation method. Then the system will represent on the tablet screen the extracted visual elements as symbols using a graphical language (this language is under-development). The third phase is to browse these graphical symbols depending on size of the used touched-screen device, and then in the fourth phase, our system provides a vibro-tactile feedback when the blind user touches the tablet. The intensity and the frequency of the vibration depend mainly on gray level under the finger. A tablet (Asus Model TF101 with Android operating system) has being used for our tests.

In this paper, we focus only on the fourth phase which specializes in giving the user a vibro-tactile feedback by transforming light contrasts of touch-screen devices into low-frequencies tactile vibrations.

To achieve the desired system, we have designed an electronic circuit which controls two micro-vibrators placed on the hands. A Bluetooth connection with an android tablet allows controlling the vibration intensity (Amplitude) of vibrators. An Android dedicated program on the tablet displays an image on the screen and detects where the user touches the tablet screen. The gray level of touched points is transmitted to the embedded device in order to control the vibration intensity. At the moment, only one micro-vibrator was used for pre-tests described in this paper.

Figure 3 illustrates the designed electronic circuit, and the used vibrator.

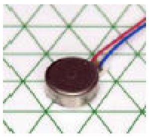


Figure 3. (a) The used micro-vibrator

Figure 3. (b) The embedded system

Figure 3. (a) The used micro-vibrator: the range of vibration frequency goes from 20 Hz up to 260 Hz. (b). The embedded system designed in the lab with the Bluetooth module.

4. PRE-TESTS PROTOCOL

4.1 Objectives of Pre-tests

Our objective of the designed protocol is enhancing the ability of VIP to recognize the 2-D structure of a web page. In order to test the prototype mentioned in section 3, we designed some images contain different structures (detailed in section 4.2), and we tested the prototype firstly on 15 sighted persons (their eyes were closed) [13], and later on 5 blind persons. Testing the protocol on sighted and blind persons gave us a more understanding of tactics and strategies achieved by sighted and blind persons to navigate the designed structures. This will be useful in designing the desired graphical vibro-tactile language (all results are detailed in next sections).

4.2 Designed Protocol for Vibro-Tactile Access

Each experiment (either for sighted or blind persons) consists of 4 ordered phases of training (learning task), and four ordered phases of evaluation (evaluation task). All the experiments were filmed, and the designed program stocked many parameters in log files (coordinates X, Y, pressure on the screen, and the time at each touch). Figure 4 presents the 4 images of training phases, and figure 5 presents the 4 images of evaluation phases.

Figure 4 (a). Image a (Training task).

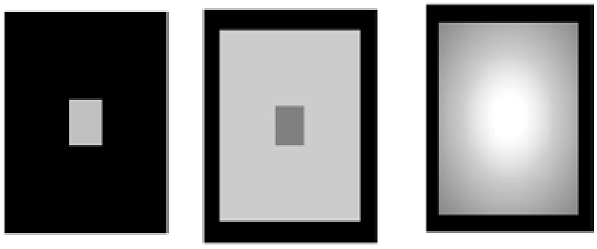


Figure 4 (b). Images b (NT2), c (NT3), d (NTG) (Training task)

Figure 4. Images of training task.

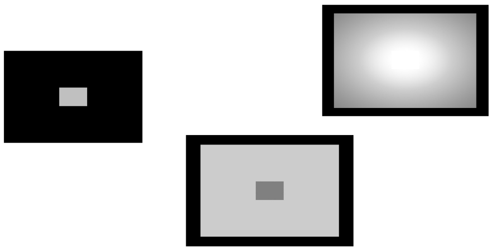


Figure 5 (a). Image a (Evaluation task).

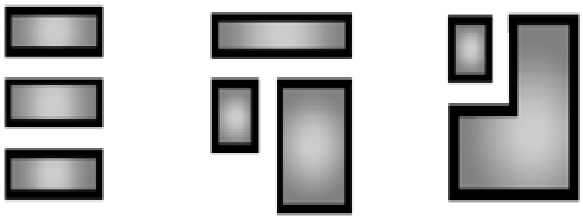


Figure 5 (b). Images b (IDP1), c (IDP2), d (IDP3). (Evaluation Task).

Figure 5. Images of evaluation tasks.

In the training task, each user discovered firstly graphical elements in each image presented in figure 4 (images a, b (NT2), c (NT3), d (NTG)), and users were informed about names of graphical elements. The name of each image NT2, NT3, NTG, indicates how many transitions are necessary to access the square center, for example NT2 proposes 2 transitions to access the center of the square.

The evaluation task consists also of 4 phases, the first one allows to discover the image 5.a and to name each square inside it, then next phases are about images 5.b, 5.c, and 5.d, where we asked users to discover contents of each image, then to describe these contents, and to redraw discovered elements inside each image. We chose these images depending on following considerations:

- Image 5.a contains all squares on which users have trained in the training task, so it could test the ability to memorize and to distinguish the shapes.

- Image 5.b contains 3 rectangles with matched sizes and with vertical order, and the image 5.c contains 3 rectangles with different sizes and many relations of directions, so testing images 5.b, and 5.c could test the ability of distinguishing sizes, and distinguishing relations of directions.

- Image 5.d contains different shapes (a rectangle and a polygon), so it could test the ability to distinguish different shapes in the same image.

- The tested images contain examples of expected results of the segmentation process, so success of distinguishing these shapes by blind users could be an indicator of their ability to distinguish results of segmenting web pages.

The results of pretests with sighted persons were already published in [1]. Table 1 presents some results of the experiment for images NT2, NT3, NTG (the time required to distinguish graphical elements and the number of errors for the 15 sighted persons. The users have been asked to name the shapes in figure 5.a, and for each shape, we evaluated the number of correct and incorrect answers). In table 1, number or errors represents the number incorrect answers.

Table 1. Results of experiments of sighted persons for images NT2, NT3, NTG. (Table extracted from [1])

Shape Name	Average Time in Seconds	Number of errors
NT2	28	2
NT3	36	6
NTG	22	3

We notice from table 1 that the lowest number of errors is assigned to image NT2, and the largest time and max number of errors is assigned to image NT3.

5. RESULTS OF EXPERIMENTS WITH BLIND PERSONS

5.1. Experiment steps:

The test performed with each one of the 5 blind persons consisted of following: personal and technical questions, explanations of the test objective, a training task, and finally an evaluation task. The approximated average time for the test for each person is about 1 hour.

5.1.1 Personal and technical questions

Before starting the tests with the 5 blind persons, we asked them to support us with some information about their age and date of their blindness. Table 2 summarizes answers of personal questions.

Table 2. Personal information of the blind persons

User-ID	0	1	2	3	4
Age (Years)	63	67	59	56	36
Sex	Male	Female	Male	Female	Female
Date of the blindness	Since Birth	since 32 years	since 25 years	Since 10 years	Since 15 years

We also asked users to provide us with some technical information about their experience in dealing with operating systems, screen readers, and what are the main problems when they navigate the Web. Table 3 shows a summary of answers for these technical questions. The first two columns indicate the number of operating systems (either Windows or Linux) used either on fixed or portable computers. The third, fourth, and fifth columns indicate the number of users who use JAWS (Job Access With Speech), NVDA (NonVisual Desktop Access), and ORCA, either on fixed or portable computers.

Table 3. Used operating systems and screen readers

	Windows	Linux	JAWS	NVDA	ORCA
Number of users with fixed computer	4	0	4	2	0
Number of users with portable computer	2	1	2	1	1

No one of the five blind persons uses a tablet, and the screen readers used with cellular phones are Talks and MobileSpeak with Nokia, and Voiceover with IPhone. Only one of the 5 persons uses a telephone to access the Web (access via IPhone). The main problems of accessing the Web via fixed or portable computers, or via IPhone telephone were: problems of access to Flash files, problems of AJAX technologies, and no ability to know the global structure of web sites. (These problems have been reported to us by the 5 users).

5.1.2 Explaining the objective of the test

To give the blind persons a good idea about the test, we explained in details what are the objectives and the phases of each tasks, and described contents of the embedded system; we also explained the final objective of the project, and why we concentrate on vibro-tactile technique regardless of other techniques.

This phase was important to initiate users for accepting kindly the test and for doing their best to interact with next steps as correctly as possible.

5.1.3 Training and learning task

In this training task, the user discovers the graphical elements in each image presented in figure 4 (images a, b (NT2), c (NT3), and d (NTG)), and the users were informed of each shape name. This task was very important for users to test the system before the evaluation task, and to know exactly how the system transforms different the grey level under the touched points on the tablet screen to a vibration mode. It is also very useful for users to control their speed of mapping the screen either to discover either the borders or the contents. During this task, the program recorded the touching information in log files (X, Y coordinates, Pressure, and Time).

Table 4 indicates training times in minutes for each user, and for each image in figure4.

Table 4. Times of training task for each user (in minutes)

User ID / Image	ID0	ID1	ID2	ID3	ID4	Total	Average
A	4,99	2,60	3,55	9,97	3,96	25,08	5,02
b (NT2)	4,40	3,11	0,99	3,30	1,00	12,80	2,56
c (NT3)	2,81	6,29	2,54	2,85	1,15	15,65	3,13
d (NTG)	3,11	3,98	2,02	3,22	1,32	13,65	2,73
Total	15,31	15,99	9,10	19,34	7,43		

We notice from table 4 that discovering the first image takes more time, and it is normal because it is the first experiment for blind users on this prototype. We notice also that there is a significant decrease in time between discovering the first and the last image in the training task. This could be an indicator that training users could decrease the time for discovering graphical elements. We can also notice the significant difference between different tested persons, for example user with ID4 needed 7.43 minutes to scan the images (A, NT2, NT3, and NTG), but the user with ID3 needed 19.34 minutes to scan the same images.

5.1.4 Evaluation task

In this task, firstly we asked each user to discover the image 5.a and to find how many squares inside it and to name each founded square, then we asked them to discover images 5.b (IDP1), 5.c (IDP2), 5.d (IDP3), and to describe them to us, and to redraw discovered shapes. Table 5 illustrates an evaluation of answers for the first question to name squares in image 5.a (The blind users have been asked to name the shapes in image 5.a, and for each shape, we evaluated the number of correct and incorrect answers). (In tables 5 and 6, the symbol ✓ represents a correct answer for the touched shape, and the symbol X represents an incorrect answer or inability to select the name of the touched shape).

Table 5. Results of questions for squares in image 4.a

User-ID/ Square Name	ID 0	ID 1	ID 2	ID 3	ID 4	Number of errors
NT2	✓	✓	✓	X	✓	1
NT3	✓	✓	X	X	✓	2
NTG	✓	X	✓	X	✓	2

We notice from table 5 that the lowest number of errors is assigned to image NT2, and it is the same result which we obtained during tests with sighted persons.

Results of answers for other questions related to images IDP1, IDP2, and IDP3 are summarized in table 6.

Table 6. Answers of questions for images IDP1, IDP2, IDP3

User-ID	ID0			ID1			ID2			ID3			ID4		
IDP	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Answers about number of rectangles	✓	✓	X	✓	X	X	✓	X	✓	X	✓	✓	✓	✓	✓
Answers about sizes of rectangles	X	X	X	X	X	X	✓	X	X	X	✓	✓	✓	✓	✓

We notice from data in tables 4, 5, and 6, that the best performance is for the user with ID4, and this may be because that this female user is the youngest between others, and it could be because that she was the only one that has already used touched devices (an IPhone in her case working with VoiceOver).

After answering questions about each image of images (IDP1, IDP2, IDP3), we asked each user to redraw the graphical elements founded in each touched image. Figure 6 views the redrawing results of the user ID4 (ID4 is the female user who gave best answers).

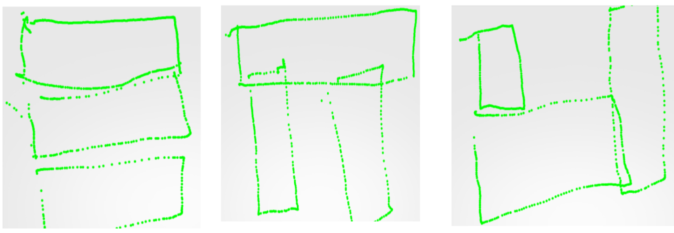


Figure 6. Results of redrawing images IDP1, IDP2, IDP3 for the user ID4.

When comparing results of redrawing (Figure 6) with images IDP1, IDP2, and IDP3, we find that the results are interesting, and we can conclude the following:

1. An ability of distinguishing sizes of shapes, because the degree of scaling between redrawn shapes is nearly equal to the degree of scaling between real shapes (IDP1, IDP2, IDP3).
2. An ability of distinguishing relations of directions, because relations of directions (vertical order, left to, right to,…) between redrawn shapes is nearly equal to relations of directions between real shapes.

The average of times in minutes consumed for each evaluation question is summarized in table 7.

Table 7. Times of the evaluation task for each user (in minutes)

User ID / Image	ID0	ID1	ID2	ID3	ID4	Total	Average
A	1,23	9,87	5,39	7,84	1,85	26,17	5,23
IDP1	4,39	14,99	1,41	3,75	1,41	25,96	5,19
IDP2	7,70	9,22	0,79	1,99	13,85	33,55	6,71
IDP3	2,71	12,94	2,81	4,03	12,58	35,06	7,01
Total	16,03	47,02	10,39	17,61	29,68		

5.2 Results Analysis

To get an idea about the most touched and the least touched areas on the screen during learning and evaluation tasks, we divided the touched-screen into 16 areas (as in figure 7, r00...r03, r10...r13, r20...r23, r30...r33), and calculated the average of touches in each area for all users.

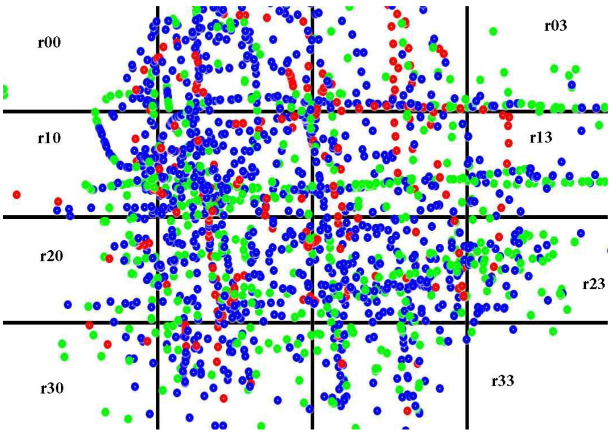


Figure 7. The 16 areas of the touched-screen of user with ID4 (Red points represent touched points with max pressure values, blue points represent touched points with pressure values less than the max and greater than the average, green points represent touched points with pressure values less than the average).

We have founded that the most touched areas are r12, r11, r22, r21, and the least touched areas are r30, r33, r32, r00 as described in figure 8. This information could be useful in next phases of our research in completing the graphical vibro-tactile language by putting the important information in the most touched areas.

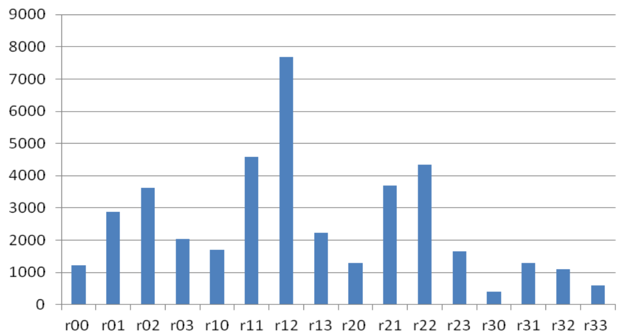


Figure 8. Most and least touched areas on the touched-screen.

During analysis the results, we have noticed that there are a lot of differences between the pressure values for all users (Pressure value depends on the used tablet; in these experiments we have used Asus Model TF101 with Android operating system).

To analyze pressure values, we calculated the max pressure value between all users, it was 3.19, and the average was 1.73, then we redrew the touched points for each user with considering that points with pressure values equal to the max value have been drawn in red color, points with pressure values less than the max and greater than the average have been drawn in blue color, and points with pressure values less than the average have been drawn in green color. Figure 7 represents an example of these points in different colors (for IDP3 evaluation task of the user who has id 4). After analyzing all the images drawn for all users, we have noticed that majority of red points (max pressure) are in images for which users gave right answers. This notice may be useful in designing our graphical vibro-tactile language, since we can observe when the user decreases or increases his touch pressure. The increasing of pressure may indicate that the user touches graphical elements interesting for him, and the decreasing may indicate that the user touches graphical elements non-interesting for him.

During the tests we observed also that users try sometimes to scan the screen very quickly, it might be because they try to get a lot of information in a short time.

6. CONCLUSION AND PERSPECTIVES

In this paper, we summarized our current work which aims to design an approach for non-visual access to web pages on touch-screen devices. The designed vibro-tactile protocol transforms the information viewed on the screen and touched by users to vibration by transforming light contrasts of touched pixels into low-frequencies tactile vibrations.

The obtained results are interesting, since we used in these experiments only one vibration motor of low level quality (phone vibrator), and the learning period was very short, so there are many enhancements to be achieved in next versions either on the hardware/software level or on the level of learning phase (increasing the number and quality of micro-vibrators, making more control on frequencies and amplitudes sent to micro-vibrators, adding vocal abilities to the current approach, integrating automatic intelligent methods for segmenting web pages,…).

In the same way that the environment enables a blind person to move in space with sidewalks and textures which will be explored by his/her white cane, we hope giving the blind user an ability to navigate documents depending on "textual sidewalks" and "graphical paths" which will be discovered by his/her finger.

Next steps in this research will be 1) Adding elements to the graphical vibro-tactile language in order to represent more HTML elements such links, buttons, input fields, and other elements, 2) Making the program more interactive to guarantee vibrating in real time, and without any delay. 3) Including the results obtained in these pre-tests, for example focusing on areas which are most touched by the users, focusing on the variation of pressure for expecting the model of navigation, 4) adding more vibration motors to the designed circuit; current version includes 2 vibration motors, but we tested only one to know exactly how blind users navigate the screen using one finger; next tests may be on more vibration motors to discover how the blind users navigate the screen using more than one finger, 5) We plan also to add thermic actuators for translating the notion of colors. This may be very useful and hopeful for blind users to transfer information about colors, 6) after adding all mentioned desired changes to the circuit and to the program; we should test integrating the hybrid segmentation algorithm of web pages with the adapted version of the designed circuit to generate automatically the graphical elements of the navigated web page.

Acknowledgment

This works is founded by the national agency of research ANR (Agence Nationale de Recherche http://www.agence-nationale-recherche.fr/) as a part of ART-ADN project ANR-12-SOIN-0003-02 (Accès par Retour Tactilo-oral Aux Documents Numériques) in GREYC laboratory (www.greyc.fr) of the University of Caen Basse-Normandie (www.unicaen.fr) and the National Superior Engineering School of Caen (www.ensicaen.fr).

Thanks to Rabeb BEN SASSI, and Abdelmajid TOUNSI for their participation in designing the system.

7. REFERENCES

[1] Maurel, F., Dias, G., Routoure, J-M., Vautier, M., Beust, P., Molina, M., Sann, C., 2012. *Haptic Perception of Document Structure for Visually Impaired People on Handled Devices*, Procedia Computer Science, Volume 14, 2012, Pages 319-329, ISSN : 1877-0509.
DOI=http://dx.doi.org/10.1016/j.procs.2012.10.036

[2] Maurel, F., Vigouroux, N., Raynal, M., Oriola, B., 2003. *Contribution of the Transmodality Concept to Improve Web Accessibility.* In Assistive Technology Research Series, Volume 12, 2003, Pages 186-193. International conference; 1st, Smart homes and health telematics; Independent living for persons with disabilities and elderly people. ISSN : 1383-813X.

[3] http://www.chromevox.com/ [Access 7/5/2014]

[4] http://www.synapseadaptive.com/gw/wineyes.htm [Access 7/5/2014]

[5] http://www.freedomscientific.com/ [Access 7/5/2014]

[6]https://play.google.com/store/apps/details?id=es.codefactory.android.app.ma.vocalizerfrfdemo&hl=fr [Access 7/5/2014]

[7]https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=fr [Access 7/5/2014]

[8] http://www.apple.com/fr/accessibility/ [Access 7/5/2014]

[9] Alaeldin, A., Mustafa , Y., Sharief, B., 2012. *Tactile WebNavigator Device for Blind and Visually Impaired People*. In Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies, Jordan, 2012.

DOI=http://dx.doi.org/10.1109/AEECT.2011.6132519

[10] BouIssa , Y., Mojahid, M., Oriola, B., Vigouroux, N., 2009. *Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents*. IEEE International Conference on Advances in Computational Tools for Engineering Applications, 2009, Pages: 591 – 594.

DOI=http://dx.doi.org/10.1109/ACTEA.2009.5227855

[11] Lenay, C., Gapenne, O., Hanneton, S., Marque, C., Genouëlle, C., *Sensory Substitution, Limits and Perspectives*. In Touch for Knowing Cognitive psychology of haptic manual perception, Amsterdam, 2003, Pages: 275-292.

[12] Tixier, M., Lenay, C., Le-Bihan, G., Gapenne, O., Aubert, D., *Designing Interactive Content with Blind Users for a Perceptual Supplementation System,* TEI 2013, 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain, 2013, Pages 229-236.
DOI= http://dx.doi.org/10.1145/2460625.2460663

[13] MAUREL, F., SAFI, W., BEUST, P., ROUTOURE, J.M., 2013. *Navigation aveugle sur dispositifs mobiles : toucher le Web... pour mieux l'entendre*, 16ème Colloque International sur le Document Électronique, CIDE16, Lille France, Europia productions, 2013.

[14] Rotard, M., Knödler, S., Ertl, T., 2005. *A Tactile Web Browser for the Visually Disabled*. In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, 2005, pages 15-22.

DOI= http://dx.doi.org/10.1145/1083356.1083361

Supervised Segmentation of Web Pages for Vibro-Tactile Access on Touch-Screen Devices

Waseem Safi Fabrice Maurel Jean-Marc Routoure Pierre Beust Gaël Dias
University of Caen Basse-Normandie
GREYC CNRS UMR 6072
Bd Maréchal Juin
14032 Caen Cedex - France
firstName.lastName@unicaen.fr

Abstract

Perception the visual layout of a web page and analyzing its DOM structure is a fundamental step in automatic adaptation process of web pages. In this paper, we present a new hybrid web page segmentation algorithm dedicated to vibro-tactile access on touch-screen devices. This hybrid algorithm combines three approaches for adapting web page contents to be presented on touch-screen devices, and for testing new navigation paradigm. The proposed algorithm is essential in our framework which aims to enhance the ability of Visually Impaired Persons (VIP) to navigate the Web by converting web pages into vibrating pages using a graphical vibro-tactile language. We present a comparison between automatic segmented pages (obtained by our proposed algorithm) and manual segmented pages. The objectives of this comparison are, on the one hand, to know how users understand web layout structures based on their visual perception, and on the other hand, to explore the main differences between automatic and manual segmentation.

I. Introduction

VIP depend on screen readers in order to deal with computer operating systems and computational programs. One of most important and desired targets by VIP is navigating the Web, considering the increased importance and expansion of web-based computational programs. Screen readers present some solutions to navigate the textual and graphical contents of web pages, either by transforming a web page into a written Braille, or by transforming the web page into a vocal output. In addition to these solutions, some screen readers installed on touch devices transform a tactile input into a vocal output.

There are some drawbacks for these proposed solutions. On the one hand, Braille techniques are costly and only few number of VIP has learned Braille [1]. On the other hand, transforming the information of a web page into a vocal format might not be suitable in public and noisy environments, and most of Braille solutions are not suitable for mobile devices [1]. In addition to these drawbacks, an important drawback is the failure to transform the 2-dimension web page structure. Indeed, as reported by many authors, perceiving the 2D structure of web documents greatly improves navigation efficiency and memorization as it allows high level text reading strategies such as: rapid or cursory reading (skimming process), finding or locating information (scanning process), to name but a few [2]. Because of these difficulties to learn and retain web site organization, most of blind people are not able to really have a "surfing experience"; Their web usage is limited to few sites they are familiar with.

Our work focuses on developing and evaluating a sensory substitution system based on a vibro-tactile solution, which may improve the mentioned drawbacks. In particular, we study how to increase the VIP perception of a 2D web page structure and how to enhance their techniques to navigate the contents of web pages on touch-screen devices. The suggested solution is cheaper than Braille devices and may be efficient in noisy and public environments compared to vocal-tactile solutions.

Our contribution is three-fold: (1) designing a Tactile Vision Sensory System (TVSS) represented by an electronic circuit and an Android program in order to transform light contrasts of touch- screen devices into haptic vibrations; (2) designing an algorithm for segmenting web pages in order to support the VIP by a way which may enhance their ability to navigate the textual and graphical contents of web pages and (3) exploring how users understand web layout structures based on their visual perception. The paper is organized as follows. First, in section 2, we review rapidly most advanced VIP targeted technologies, and propose our own framework. In section 3, the state of the art for webpages segmentation methods is reviewed. In the fourth section, our hybrid segmentation method is presented. In section 5, we present the results obtained in an experiment where the differences between automatic and manual web page segmentation are compared. Finally, perspectives and conclusions are presented.

II. VIP TARGETED TECHNOLOGIES AND PROPOSED FRAMEWORK

Current products for VIP such as screen readers mainly depend on speech synthesis or Braille solutions, e.g. ChromeVox [1], Windows-Eyes [2], or JAWS (Job Access With Speech) [3]. Braille displays are complex and expensive electromechanical devices that connect to a computer and display Braille characters. Speech synthesis engines convert texts into artificial speech, where the text is analyzed and transformed into phonemes. These phonemes are then processed using signal processing techniques.

Some screen readers can also support tactile feedback when working on touch-screen devices, such as Mobile Accessibility and Talkback [4] for Android, or VoiceOver [5] for iOS. Many of these products propose shortcuts for blind users to display a menu of HTML elements existing in the web page, for example headers, links and images. But, the main drawback of all these products is the fact that they transfer the information of web pages into a linear way i.e. without any indication of the 2-dimension global structure.

Many researches tried to enhance the way VIP interacts with web pages, such as [3], who proposed a tactile web navigator. This navigator extracts texts from web pages and sends them to a microcontroller responsible of displaying the text in Braille language using an array of solenoids. A tactile web browser for hypertext documents has been proposed by [4]. This browser renders texts and graphics for VIP on a tactile graphical display and supports also a voice output to read textual paragraphs and to provide a vocal feedback. The authors implemented two exploration modes, one for bitmap graphics and another one for scalable vector graphics. A pin matrix device is used to produce the output signal for blind users. The main drawback of these two proposed systems is that they need specific devices (solenoids and pin matrix), which are expensive and cannot be integrated in handled devices such as PDAs or Tablet PCs. Another interesting model called MAP-RDF (Model of Architecture of Web Pages) has been proposed by [5]. This model allows representing the structure of a web page and provides blind users an overview of the web page layout and the document structure semantics. Finally, Tactos is a perceptual interaction system, which has been suggested by [6] and consists of three elements: (1) tactile simulators (two Braille cells with 8 pins) represent a tactile feedback system, (2) a graphics tablet with a stylus represents an input device and (3) the computer. More than 30 prototypes of Tactos have been released to serve a lot of users in many domains. Tactos has been successfully used to recognize simple and complex shapes. The device has been also used in geometry teaching domain in an institution for visually impaired and blind children. Tactos also allowed psychology researchers to propose and develop new paradigms for studying perceptions and mediated communication of blind persons [7]. However, it shows the same drawbacks (expensive and need specific devices) as the previous systems. Moreover, the blind user can only explore the web page with a stylus and both hands are occupied by the system. Moreover, it is unemployable for a large set of environments, for example in public.

The "first glance" can be defined as the ability to understand the document layout and its structural semantics in a blink of an eye [3]. In this work, we aim to increase the ability of VIP to understand the 2-dimension web page layout in order to enhance their tactics to navigate the Web with a vibro-tactile feedback. The first phase in our model is to extract visual structures in the navigated web page and convert these "visual" blocks into zones (or segments) to facilitate the navigation in later phases. We achieve this phase using a hybrid segmentation method. Then the system represents the extracted visual elements as symbols using a graphical language. The third phase is to browse these graphical symbols depending on the size of the used touched-screen device; and in the fourth phase, our system provides a vibro-tactile feedback when the blind user touches the tablet by giving the user a vibro-tactile feedback by transforming light contrasts of touch screen devices into haptic vibrations.

To achieve the desired system, we have designed an electronic circuit, which controls four micro-vibrators placed on two fingers. A bluetooth connection with an Android tablet allows controlling the vibration intensity (i.e. amplitude) of vibrators. An Android dedicated program on the tablet displays an image on the screen and detects where the user touches the tablet screen (the viewed image represents the result of web page segmentation). The light intensity at touched points is transmitted to the embedded device in order to control the vibration intensity. In this paper, we focus only on the first phase (extracting visual structures in the navigated web page, and converting them into zones). A detailed description of hardware components of the system and results of pre-tests are described in [2] [8], and [18].

III. RELATED WORKS

Segmenting a web page is a fundamental phase for understanding its global structure. Extracting the global structure of web pages is useful in many domains such as information retrieval, data extraction, and similarity of web pages. Many approaches

[1] http://www.chromevox.com/ [Access 10/12/2014]

[2] http://www.synapseadaptive.com/gw/wineyes.htm [Access 10/12/2014]

[3] http://www.freedomscientific.com/ [Access 10/12/2014]

[4] https://play.google.com [Access 10/12/2014]

[5] http://www.apple.com/fr/accessibility/ [Access 10/12/2014]

have been suggested for segmenting web pages, such as:

a) DOM-based segmentation: it depends on analyzing the DOM tree (Document Object Model), and extracting the main structure of web pages depending on HTML tags as described in [9]. In this paper, the authors determine firstly the layout template of a web page, and then it divides the page into minimum blocks, and finally collects these minimum blocks into content blocks.

b) Vision-based segmentation: this method divides the web page depending on the visual view of web page contents on a web browser. The well known tool VIPS (VIsion based Page Segmentation) [10] is based on this approach.

c) Image processing based segmentation: this approach captures an image for the visual view of a web page, and then depends on image processing techniques to divide the captured image into sub-blocks [11] [12].

d) Text-based Segmentation: this approach focuses on extracting only information about texts in a web page. After dividing the web page into blocks of texts, it could be possible to find the semantic relations between these textual blocks. This method is useful in many information retrieval domains such as question answering applications [13].

e) Fixed-length segmentation: this approach divides the web pages into fixed length blocks (passages), after removing all HTML tags, where each passage contains a fixed number of words [14].

f) Densitométrie analysis based segmentation: this approach depends on methods applied in quantitative linguistics, where text-density refers to a measure for identifying important textual segments of a web page [15].

g) Graph-based segmentation: This approach depends on transforming the visual segments of a web page into graph nodes, then applying many common graph methods on these nodes for combining them into blocks, or for making a clustering for these nodes. Some common works which depend on this approach are [16] [17].

h) and Hybrid-based segmentation: This approach combines many approaches indicated previously.

IV. SUGGESTED HYBRID SEGMENTATION ALGORITHM

Most of segmentation algorithms render first the web page using a web browser, and then segments the HTML elements into many blocks depending on the visual layout. Our constructed hybrid segmentation algorithm has been tested on 154 pages collected manually from many newspapers and e-commerce sites (www.leparisien.fr, www.lefigaro.fr, www.liberation.fr, www.amazon.fr, www.materiel.net, wwww.photobox.fr). The results have been integrated with our under-development Android program. The obtained results are promising because the segmentation algorithm can efficiently extract the web page blocks depending on the visual structure, and the algorithm can also convert correctly these blocks into zones (clustering the blocks). Our algorithm mixes three segmentation approaches, DOM-based segmentation, vision-based segmentation, and graph-based segmentation.

A. Vision-Based Approach:

In this phase, we render the web page using Selenium [6] web driver and Mozilla FireFox browser, and we get its visual structure by Java-script code injection inside the HTML source code of the rendered web page. The obtained visual structure indicates a global hierarchy of the rendered web page. This phase assigns additional information for each DOM HTML element such as XPath and bounding box (location [X0, Y0], and size [height and width]). The input of this phase is a web page HTML source code, and its output is an augmented HTML web page with injected infor- mation about bounding boxes and DOM XPath for each HTML elements. In next sections, we refer to bounding boxes by blocks (i.e. each bounding box represents an HTML element, and may contain other bounding boxes).

The main steps to obtain these blocks for a web page W are:

```
1. Running Selenium web driver (rendering
   the web page W using FireFox browser),
2. Getting the DOM-structure elements
   DOM-Tree:=DOM(W),
3. For each element e (Block) in DOM-Tree
       elements:
     3.1. Get information about element e (XPath,
          and bounding box),
     3.2. Inject the obtained information in HTML
          source,
4. Return the augmented web page (DOM-Tree with
   injected information).
```

[6] http://www.seleniumhq.org/

B. DOM-Based Approach:

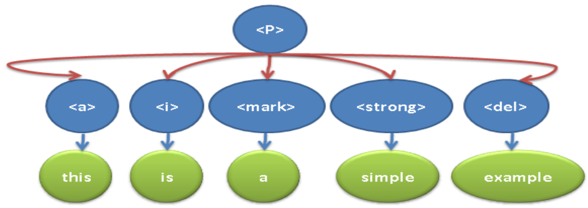
After obtaining the visual structure of a web page, we analyze its DOM structure by applying filters and re-organization rules for enhancing results of next phases. We divide the DOM elements depending on the specification of HTML5 content models proposed by the World Wide Web Consortium (W3C) [7]. This specification divides the HTML tags to 7 categories (Metadata content, Flow content, Sectioning content, Heading Content, Phrasing content, Embedded content, and Interactive content). The first applied filter is Metadata-Content-Filter, which deletes all the elements considered as metadata content elements except "title" tag. We delete the other tags because they do not contain useful visual information in next steps. The deleted tags are "base", "command", "link", "meta", "noscript", "script", and "style". We do not delete the "title" node because the textual information existing in this node might be useful for the user (the page title). We add also to this filtered group the following tags "comment", "br", and "doctype".

The second applied filter is Dead-Nodes-Filter, where it deletes all HTML nodes that do not affect on the appearance, for example nodes with height or width equals to "0px" (zero pixel); or nodes with style properties ("display:none" or "visibility:hidden" or "hid- den:true"). After applying the previous filters, we apply some reorganization rules in order to enhance visualizing the information in next phases. One example of these rules is Paragraph-Reorganization-Rule, where this rule re-constructs all paragraph child-nodes in one node which should contain the extracted sub-texts. We made this rule after analyzing many DOM structures, and observing that the text in some paragraph nodes is distributed between many child-nodes such as ⟨i ⟩ (italic), ⟨span⟩, ⟨b⟩ (bold), ⟨strong⟩, ⟨em⟩ (emphasized), ⟨small⟩, ⟨mark⟩(marked), ⟨del⟩ (deleted), ⟨ins⟩ (inserted), ⟨cite⟩ (defining a title of work), ⟨u⟩ (underline), and ⟨sub⟩ (subscript). So extracting these subtexts and collecting them in one text is useful and more efficient for visualizing them as one block in next phases rather than visualizing them as many separated blocks.

To illustrate the result of applying this rule, let us consider the following HTML source code:

```
<p>
    <a> this </a>
    <i> is </i>
    <mark> a </mark>
    <strong> simple </mark>
    <del> example. </del>
</p>
```

Its DOM structure is:



And after applying the rule, the DOM representation will be:

This rule is applied on all tags of type ⟨P⟩ (paragraph), and on all tags existed in the heading content category (specification of HTML5 content models), this group contains the following nodes (h1, h2, h3, h4, h5, h6, and hgroup). We also apply this rule on many other HTML tags which might contain textual child nodes such as ⟨a⟩ (hyperlink tag), ⟨abbr⟩ (abbreviation), ⟨acronym⟩ (this tag is not supported in HTML5), ⟨address⟩ (contact information for the author/owner), ⟨bdi⟩ (Bi- directional Isolation), ⟨button⟩, ⟨label⟩, ⟨li⟩ (list element), and ⟨q⟩ (quotation). We used Jsoup tool [8] in order to access to a web page DOM structure, and getting its HTML hierarchy. The result of this phase is a filtered DOM-tree; each of its nodes is visible and contains XPath and bounding box information. The designed filters and re-organization rules were integrated with our framework, and then we applied these rules and filters on the vision-based segmented web pages (154 pages mentioned pre-viously). After getting the filtered DOM-tree for each page, we represented the obtained bounding boxes on the used tablet Sam sung GALAXY Tab 2 (10.1inch, dimensions HeightXWidthXDepth 175.3X256.7X9.7 mm, Android version 4.1.2), after making a scaling for sizes of bounding boxes to be appropriate with the new size of used tablet.

C. Graph-Based Approach:

After segmenting the web page depending on its visual structures and analyzing its DOM-structure, we apply a new graph-based segmentation algorithm called "Blocks2Zones Clustering" in order to group many similar blocks together in one zone. Clustering many blocks together is necessary in order to decrease the number of viewed blocks in some interfaces (instead of viewing many blocks, we view one zone which represents these blocks and then the user can navigate intra-elements inside the zone by double clicking on the graphical element of the chosen zone), and to group closed blocks in one zone (here, closeness depends on distances between blocks, this will be described next sections in details). The pseudo-code of the proposed algorithm is:

[7] http://www.w3.org/TR/2011/WD-html5-20110525/content-models.html

[8] http://jsoup.org/

```
Blocks2Zones Clustering Algorithm
Input (Blocks, Number of desired Zones)
Output: Graph of N nodes (N Zones)
1- Transform the blocks into a graph
 (Undirected graph)
     1.1. Blocks --> Nodes,
     1.2. Make relations between the nodes,
     and assign weights for these relations.
2- If number of zones <= number of blocks
            end the algorithm,
 Else
 3- Find the smallest size node (node A)
 4- For node A, find the connection which has
    the largest weight (node B).
 5- Group the nodes A, and B (A+B).
 6- Repeat steps 3,4,5 till number of blocks
     ==  number of zones.
```

We can describe the output graph as following G=(V,E), where G is undirected graph, V is a set of vertices (nodes or zones), and E is a set of edges (connections between zones), and |V|=n (number of desired zones). We define the set of vertices V={v_i: 1≤i≤n where n is the number of desired zones, and v_i is set of sub-zones}. We define the set of connections as following E={e_j: $e_j(v_{j1},v_{j2})$: v_{j1} and v_{j2} are nodes in V}. To calculate the weights between nodes, we tested 2 relations of distances: the first one is Manhattan distance (d(p,q) = d(q,p)= $\parallel p-q \parallel = \sum_{i=1}^{n}$ |p_i-q_i |, and the second is Euclidian distance (d(p,q) = d(q,p)= $\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+.....+(p_n-q_n)^2}$)). Where p and q are centers of two nodes between them we want to calculate the distance. To ensure which distance should be used, we applied an internal quality criterion for the two used distance relations; the applied criterion is Sum of Squared Error (SSE) (SSE = $\sum_{K=1}^{n}\sum_{\forall X_i \in C_k}^{n} \parallel X_i - \alpha_i \parallel^2$ Where C_k is the set of instances in cluster K, and $\alpha_{k,j} = \frac{1}{N_K}\sum_{\forall x_i \in C_k}^{n} x_{i,j}$). Results of applying SSE measure on the two distances (Manhattan, and Euclidian) were identical; which means that using either Manhattan distance or Euclidian distance is equal in our algorithm to calculate weights between nodes.

In following, an example of applying this algorithm on the main page of web site w3schools.com is given. In this example we want that the algorithm segments the main page into 7 zones. The first iteration of the algorithm will divide the page into 13 zones (depending on the DOM structure) as illustrated in figure 1.a, and will construct a graph of 13 nodes as illustrated in figure 1.b.

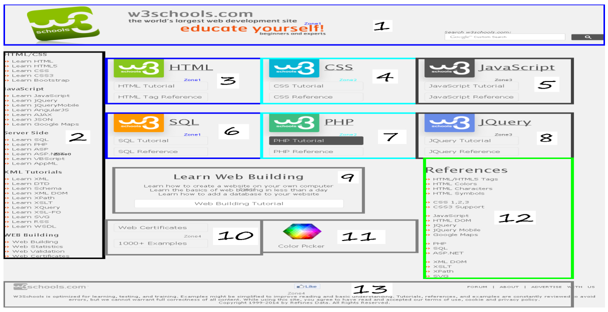


Figure 1.a Segmentation for 13 zones

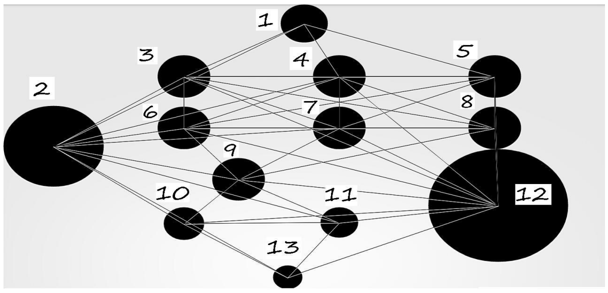
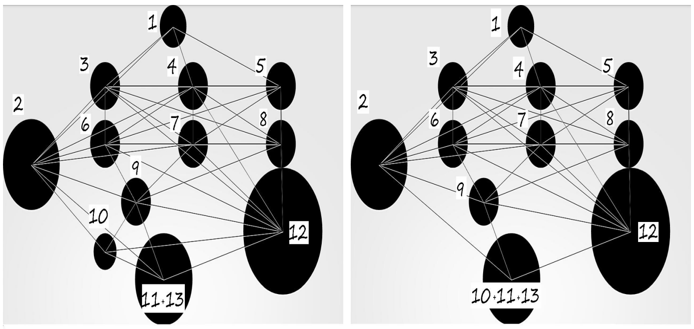
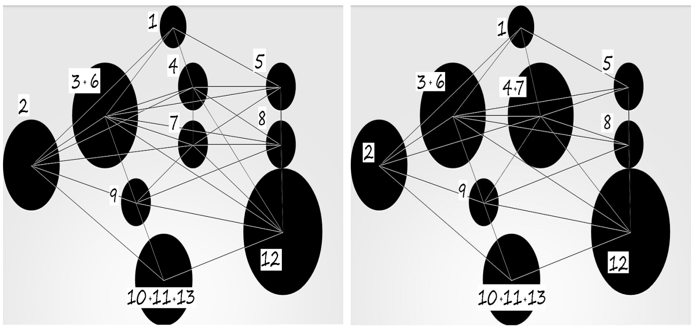


Figure 1.b Constructed graph for 13 zones
Figure1: First segmentation of www.w3schools.com

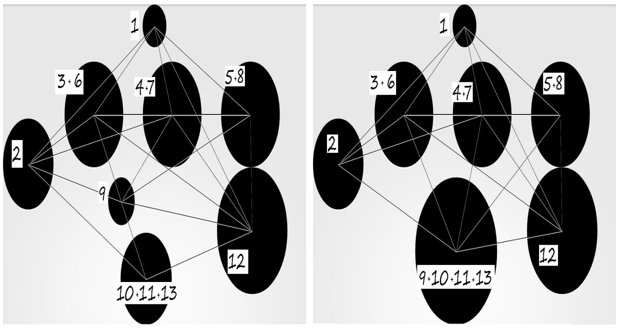
Figures 2, 3, and 4, illustrate next iterations of the algorithm and how the 13 zones are converted to 7 zones.



2.a 12 zones segmentation 2.b 11 zones
Figure2. 12, 11 zones segmentation of site www.w3schools.com



3.a 10 zones segmentation 3.b 9 zones
Figure3. 10, 9 zones segmentation of site www.w3schools.com



4.a 8 zones segmentation 4.b 7 zones
Figure4. 8, 7 zones segmentation of site www.w3schools.com

Applying this hybrid segmentation algorithm on a filtered DOM-tree (obtained from applying Vision-based approach then DOM-based approach) converts a web page to a set of zones, each zone contains many other zones or blocks, since each block represents a visual structure of HTML element and may contain many other blocks. The purpose of the proposed vibro-tactile access protocol is then to transform semantics of symbols in these zones, or blocks, or HTML elements into vibrations with different frequencies and amplitudes.

V. MANUAL AND AUTOMATIC WEB PAGE SEGMENTATION DIFFERENCES

In order to evaluate our algorithm results, and to know how users understand web layout structure based on their visual perception, we made an experiment: we asked 15 volunteers to make a manual segmentation of different kinds of web pages. The volunteers are different ages (between 25 and 50 years old), and most of them are informatics specialists. We presented for each volunteer 4 printed copies (A4 size papers) of 8 web pages (2 pages from www.cdiscount.com, 2 pages from www.photobox.com, 2 pages from www.rueducommerce.fr, one page from www.w3schools.com, and one page from www.leparisien.fr). We asked each volunteer to segment the 4 copies of each web page into 3, 4, 5 and 6 zones with following considerations:

- all the pages are printed in gray scale in order to avoid affecting the colors on the segmentation process. We chose this option because the current version of our algorithm does not depend on color differences between blocks,

- users can segment the page using polygons with minimum of 4 points (triangles are not allowed),

- Intersecting between segmented polygons are not allowed,

- users can start segmenting from any part or direction of the page (left, right, top, down, or center),

- we asked users to write the order number of zones (inside or beside the zone) while they make the segmentation process; this is very useful for us to know how the users start segmenting the pages, and how they finish it.

After collecting all manual segmented copies (480 papers : 15 users X 8 websites X 4 copies), we noticed the following:

- A lot of users do not start the segmentation process for certain number of zones by the same way they start segmenting for other number of zones,

- 16% of papers have been segmented starting from the center of the page, 3.5% have been segmented starting from the bottom of the page, and the majority of papers 80.5% have been segmented starting from top of the page,

- 20% of papers have been segmented vertically, and 80 have been segmented horizontally,

- 92.6% of segments are rectangles, and 7.4% of segments are polygons with more than 4 points,

- The most similar results between users were obtained for pages segmented in 3 zones, and the most different results between users were obtained for pages segmented in 6 zones.

- Finally, we noticed that it is very difficult to detect a segmentation method common between all users; since each user segments the pages depending on his understanding of the web page layout structure, on his visual perception of the visible elements, and on his interests and visual experience.

We also run our algorithm with the 8 mentioned web pages. The algorithm segmented each web page into 3, 4, 5, and 6 zones. The comparison between manual and automatic segmentation is illustrated in figures 5 and 6. Figure 5 illustrates matching results based on what we call “Strong criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if the results are 100% identical without any difference. We can conclude from figure 5 that pages segmented automatically and manually into 4 zones are more matched than pages segmented into other numbers of zones. The percentage of identical matching depending on the strong criterion is 15.41% (74 identical matched results of 480 segmented copies).

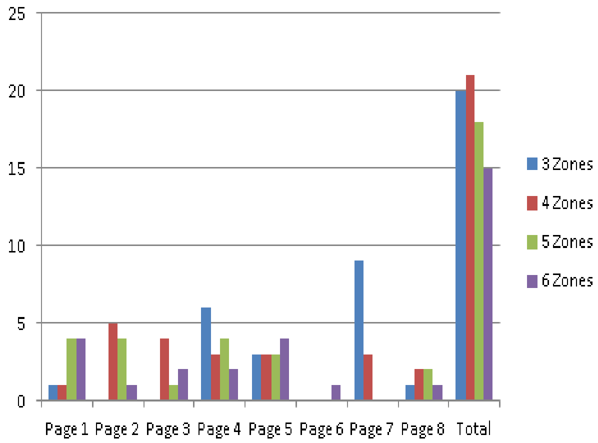


Figure5. Strong-criterion based matching results of manually and automatic segmentation

Figure 6 illustrates matching results based on what we call "Weak criterion". Depending on this criterion, we can consider that 2 segmentation results are matched if at least 50% of the results are identical.

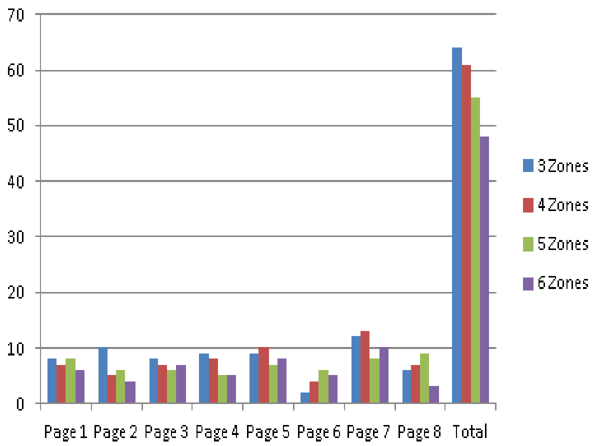


Figure6. Weak-criteria based matching results of manually and automatic segmentation

We can conclude from figure 6 that pages segmented automatically and manually into 3 zones are more matched than pages segmented into other numbers of zones. The matching percentage depending on the weak criterion is 47.5% (228 identical matched results of 480 segmented copies).

These results are promised considering that this is the first version of the algorithm, and especially that this version does not take into account the human visual perception aspects, and does not use image processing techniques which could be very useful in enhancing the achieved results.

VI. CONCLUSION AND PERSPECTIVES

In this paper, we summarized our current work which aims to design an approach for non-visual access to web pages on touch-screen devices, and we focused on the suggested hybrid segmentation algorithm. We expect that integrating this method of segmentation with the designed vibro-tactile protocol can give VIP an impression of the first glance layout of web pages.

Next steps in this research will be 1) adding some image processing techniques in order to enhance the proposed segmentation algorithm 2) adding advanced techniques in text summarization to facilitate navigating textual information, 3) making real experiments to study effects of suggested segmentation algorithm on VIP web navigation models.

References

[1] F. Maurel, G. Dias, J-M. Routoure, M. Vautier, P. Beust, M. Molina, and C. Sann, *Haptic Perception of Document Structure for Visually Impaired People on Handled Devices* , Procedia Computer Science, Volume 14, Pages 319-329, ISSN : 1877-0509, 2012. DOI=http://dx.doi.org/10.1016/j.procs.2012.10.036

[2] F. Maurel, N. Vigouroux, M. Raynal, and B. Oriola, *Contribution of the Transmodality Concept to Improve Web Accessibility*, In Assistive Technology Research Series, Volume 12, 2003, Pages 186-193. International conference; 1st, Smart homes and health telematics; Independent living for persons with disabilities and elderly people. ISSN : 1383-813X. 2003.

[3] Alaeldin, Y. Mustafa, and B. Sharief, *Tactile WebNavigator Device for Blind and Visually Impaired People*, In Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies, Jordan, 2011. DOI=http://dx.doi.org/10.1109/AEECT.2011.6132519

[4] M. Rotard, S. Knödler, and T. Ertl, *A Tactile Web Browser for the Visually Disabled*, In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, 2005, pages 15-22, 2005. DOI= http://dx.doi.org/10.1145/1083356.1083361

[5] MY. BouIssa, M. Mojahid, B. Oriola, and N. Vigouroux, *Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents*, IEEE International Conference on Advances in Computational Tools for Engineering Applications, 2009, Pages: 591 – 594, 2009. DOI=http://dx.doi.org/10.1109/ACTEA.2009.5227855

[6] C. Lenay, O. Gapenne, S. Hanneton, C. Marque, and C. Genouëlle, *Sensory Substitution, Limits and Perspectives*, In Touch for Knowing Cognitive psychology of haptic manual perception, Amsterdam, Pages: 275-292, 2003,

[7] M. Tixier, C. Lenay, G. Le-Bihan, O. Gapenne, and D. Aubert, *Designing Interactive Content*

with Blind Users for a Perceptual Supplementation System,TEI 2013, 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain, Pages 229-236, 2013. DOI= http://dx.doi.org/10.1145/2460625.2460663,

[8] F. Maurel, W. Safi, P. Beust, and J.M. Routoure, *Navigation aveugle sur dispositifs mobiles : toucher le Web... pour mieux l'entendre*,16ème Colloque International sur le Document Électronique, CIDE16, Lille, France, Europia productions, 2013.

[9] A. Sanoja, and S. Gançarski, *Block-o-Matic: a Web Page Segmentation Tool*,BDA. Nantes, France. 2013. http://hal.archives-ouvertes.fr/hal-00881693/.

[10] C. Deng, Y. Shipeng, W. Ji-Rong, and M. Wei-Ying, *VIPS: a Vision-based Page Segmentation Algorithm*,Nov. 1, 2003, Technical Report MSR-TR-2003-79, Microsoft Research. 2003. http://research.microsoft.com/pubs/70027/tr-2003-79.pdf

[11] D. Cai, X. He, W-Y. Ma, J-R. Wen, and H. Zhang, *Organizing WWW Images Based on the Analysis Of Page Layout And Web Link Structure*,Microsoft Research Asia, Beijing, China, 2004. http://research.microsoft.com/pubs/69080/25.pdf

[12] J. Cao, B. Mao, and J. Luo, *A segmentation method for web page analysis using shrinking and dividing*,International Journal of Parallel, Emergent and Distributed Systems - Network and parallel computing, Volume 25 Issue 2, April 2010. Pages:93-104, 2010. DOI=http://dx.doi.org/10.1080/17445760802429585

[13] N. Foucault, S. Rosset, and G. Adda, *Pré-segmentation de pages web et sélection de documents pertinent en Questions-Réponses*,ITALN-RÉCITAL 2013.

[14] J.P. Callan, *Passage- level Evidence in Document Retrievals*, the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Pages: 302-310. Publisher: Springer-Verlag New York, Inc, 1994. http://dl.acm.org/citation.cfm?id=188589

[15] C. Kohlschütter, and W. Nejdl, *A Densitometric Approach to Web Page Segmentation*, USA, CIKM'08, Proceedings of the 17th ACM conference on Information and knowledge management, 2008. Pages: 1173-1182. DOI: http://dx.doi.org/10.1145/1458082.1458237

[16] D. Chakrabarti, R. Kumar, and K. Punera, *A graph theoretic approach to webpage segmentation*,Proceedings of the 17th international conference on World Wide Web, WWW'08, ACM, USA, 2008. Pages: 377-386. Publisher: ACM New York, NY, USA, 2008. DOI: http://dx.doi.org/10.1145/1367497.1367549

[17] X. Liu, H. Lin, and Y. Tian, *Segmenting Webpage with Gomory-Hu Tree Based Clustering*, Journal of Software, Vol 6, No 12, Pages: 2421-2425. 2011.

[18] W. Safi, F. Maurel, J-M. Routoure, P. Beust, and G. Dia, *Blind Browsing on Hand-Held Devices: Touching the Web to Understand it Better*, Data Visualization Workshop (DataWiz2014) associated to 25th ACM Conference on Hypertext and Social Media (HYPERTEXT 2014).

Navigation aveugle sur dispositifs mobiles : toucher le Web... pour mieux l'entendre

Blind Browsing on Handheld Devices: Touching the Web… to Hear it Better

Fabrice MAUREL, Waseem SAFI, Pierre BEUST, Jean-Marc ROUTOURE

GREYC, Université de Caen Basse-Normandie, Campus Côte de Nacre
Caen, 14032, France – Soutenu par l'ANR (ART-ADN : ANR-12-SOIN-0003-02)
Fabrice.maurel@unicaen.fr

Résumé. L'accès universel à l'information du Web est un enjeu sociétal, technologique, et scientifique majeur. Pour accéder aux documents électroniques, les aveugles utilisent des lecteurs d'écran afin d'extraire l'information, la linéariser, puis la transformer en Braille grâce à du matériel spécifique et/ou l'énoncer via une synthèse vocale. Par ailleurs, nous savons que la perception de la structure visuelle des textes joue un rôle important dans la mémorisation, la compréhension et les stratégies de lecture. Afin de rendre moins linéaire et plus active la navigation non visuelle, nous proposons de transformer les contrastes lumineux de l'écran, induits par la mise en forme, en vibrations tactiles perceptibles par un équipement peu intrusif et économique. Les premières expérimentations sur des tablettes offrent des résultats prometteurs dans un paradigme « design for all ».
Mots-clés. Perception haptique, structure visuelle des documents, actionneurs basse-fréquence, dispositifs mobiles, design for all

Abstract. Web universal access is a societal, technological and scientific major issue. To access electronic documents, blind people uses screen readers in order to extract information, linearize it, either transform it to be presented in Braille with specific hardware and/or to be spoken by a voice synthesizer. Moreover, we know that text layout plays an important role in memory, comprehension and reading strategies. To make non-visual browsing less linear and more active, we propose to transform bright contrasts of the display, induced by the document layout, in tactile vibrations perceptible by a minimally intrusive and economic equipment. The first experiments on touch tablets offer promising results in a "design for all" paradigm.
Keywords. Haptic Perception, Document Structure Skeletons, Low-Frequency Actuators, Handheld Devices, Design-for-All

1 Introduction

L'OMS estime que près de 314 millions de personnes sont sujettes à diverses déficiences visuelles (dont 39 % en raison de problèmes liés à l'âge), parmi lesquelles 45 millions sont aveugles. En France, 10 % de la population connait des problèmes de vision plus ou moins forts. Une estimation précise évalue que sur 750000 naissances, 100000 ont ou auront des problèmes de vue. Aujourd'hui, la population aveugle est d'environ 1/1000, soit 77000 personnes dont 15000 seulement ont appris le Braille et 7000 le pratique régulièrement.

Avec la prolifération des nouveaux dispositifs d'accès à l'information, la nécessité d'interfaces pour accéder aux contenus multimédia apparaît clairement. Les solutions classiques utilisées par les déficients visuels (DV) pour interagir avec un ordinateur présentent certaines limites (Rotard *et al.*, 2005), (Alaeldin *et al.*, 2011), (Bou Issa *et al.*, 2009), (Ziat *et al.*, 2007), (Lenay *et al.*, 2003). Premièrement, elles sont mal adaptées aux dispositifs mobiles, encombrants et relativement chers. De plus, les environnements bruités et publics rendent difficiles l'utilisation de la synthèse ou de la reconnaissance de la parole. Enfin, des études antérieures (Pontelli *et al.*, 2002), (Maurel *et al.*, 2003) ont démontré l'importance de la perception globale des documents pour initier une interaction cognitivement efficace. Pourtant, peu de solutions parviennent à compenser de manière satisfaisante l'impossibilité pour un non-voyant de cibler rapidement et directement une partie de l'écran comme un lecteur visuel. Il est parfois fait référence à cette capacité comme un « premier regard » qui permet de prendre en compte d'un coup d'œil la mise en page des documents et donc la sémantique qui passe par la structure visuelle. Les lecteurs aveugles doivent au contraire parcourir les contenus séquentiellement, en subissant divers procédés de linéarisation automatique puisque les pages web sont des structures bidimensionnelles (Earl et Leventhal, 1999), (Gunderson et Mendelson, 1997) comme illustrées par la figure 1.

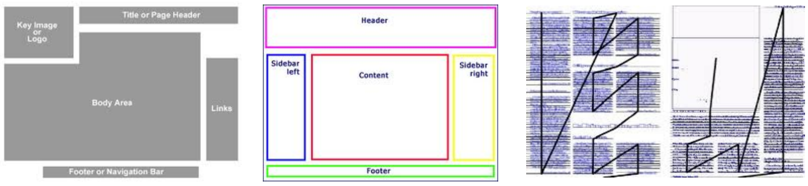


Figure 1. *Différentes mises en page de documents web*

Accéder à cette sémantique qui passe par la structure visuelle des textes, c'est accéder à toutes les stratégies de haut niveau de lecture qu'elle favorise : lire en diagonale, rechercher/retrouver rapidement une information, pointer directement une partie du texte... Notre objectif est de donner une perception non visuelle de cette « architecture » qui organise et facilite le parcours des documents et des interfaces. La solution que nous proposons consiste à exploiter la typo-disposition pour générer des pages web vibrantes ; pour cela nous transformons les contrastes lumineux produits par l'écran en vibrations tactiles basse-fréquence perceptibles via des actionneurs placés sur une partie du corps. Les documents peuvent alors être survolés sur une tablette tactile et l'information est fournie à l'actionneur qui calcule en temps réel la fréquence de vibration correspondante. Autrement dit, nous essayons avec ce dispositif de remplacer l'exploration visuelle classique d'un document (basée sur la perception des vibrations lumineuses de l'écran) par une exploration non visuelle (basée sur la perception des vibrations tactiles d'un actionneur).

Par rapport aux travaux existants, nous proposons une solution sans Braille, économique, peu intrusive, discrète, temps-réel et facilement utilisable dans le contexte de la mobilité. Les premières expérimentations sur tablettes tactiles sont prometteuses dans le paradigme « design pour tous ».

2 Contexte commercial

La synthèse de parole à partir de textes et le Braille sont les principales solutions utilisées par les non-voyants pour accéder aux ordinateurs. Ces deux technologies sont utilisées en sortie de lecteurs d'écran. Les afficheurs Braille sont des dispositifs électromécaniques assez chers qui se connectent à l'ordinateur et sont capables de donner accès en temps réel à une série de caractères sous la forme de picots Braille. Les synthèses de la parole à partir de textes (TTS) convertissent les phrases en parole artificielle. Un traitement linguistique repère et analyse les phrases qui sont ensuite transcrites en phonèmes à leur tour transformés par des techniques de traitement du signal en sons numériques.

Une alternative à cette technologie est l'utilisation de navigateurs parlants qui proposent une lecture simplifiée des informations. Depuis 2005, Opera a développé une version de son navigateur capable de lire des portions de texte sélectionnées. Firefox et Chrome offrent des extensions vocales telles que FireVox et ChromeVox. Par exemple, ces systèmes annoncent le titre de la page web avant de lire son contenu. Il est toutefois important de noter que, quelle que soit la technologie utilisée, l'utilisateur non-voyant n'aura jamais une vue globale de la structure bidimensionnelle du document ; ces solutions permettent seulement une perception linéaire de différents fragments qui doivent être manipulés mentalement afin de reconstruire l'information complète. Malgré tout, ces dernières années, fortement influencées par la politique d'accessibilité d'Apple, des solutions logicielles ont émergé pour l'accès tactilo-oral aux contenus sur écrans multitouches. Par exemple, le principal moyen d'accès non visuel à un Ipad est VoiceOver, un lecteur d'écran avec un moteur de synthèse vocale. Cette dernière solution permet d'obtenir des informations orales instantanées sur les éléments survolées d'un geste sur l'écran tactile. Un bouton virtuel de commande, appelé « rotor », permet de choisir une unité telle que titres, liens, tableaux et images ; une action peut alors sélectionner un élément et naviguer entre ses différentes occurrences. Cette technique est avantageusement couplée avec les « web spots » : ces marques, automatiquement générées par VoiceOver pour chaque page appelée par le navigateur ou placées par l'utilisateur, indiquent les différentes parties intéressantes d'une page web ; le « rotor » permet alors de choisir de suivre ce parcours de lecture. Malheureusement, nous n'avons pas trouvé de données scientifiques sur l'impact de ces techniques d'interaction non visuelles ; il est toutefois important de noter que de nombreux témoignages de non-voyants assurent une expérience positive (rapport du séminaire d'Apple sur l'intégration des personnes handicapées le 1er février 2011 à Paris).

Bien que ces contributions aillent dans la bonne direction, elles sont grandement facilitées par une architecture fermée et des interfaces entièrement contrôlées. De plus, aucune de ces solutions ne compensent l'absence du « premier regard » qui permet une intégration précoce des indices visuels utiles pour contrôler efficacement l'interaction. La section suivante décrit les approches scientifiques les plus innovantes qui travaillent à cet objectif.

3 Contexte scientifique

Le Braille permet la représentation des caractères, de la ponctuation et d'une partie de la mise en page ; Dans un paradigme de conception pour tous (Bougie *et al.*, 2001), nous évitons ce code afin de ne pas exclure les personnes voyantes et la plupart des DV qui ne le lisent pas. Quelques concepteurs de synthèse vocale ont vu la nécessité « d'aller au-delà de la limite de la phrase (le paragraphe ou même une organisation hiérarchique des documents), pour une lecture non linéaire (Boula de Mareüil *et al.*, 2002) ». Pourtant, quand on examine les systèmes récents, nous remarquons que la typo-disposition est peu utilisée dans le calcul de la prosodie de synthèse. Une proposition a été faite afin de transposer l'architecture visuelle des documents vers la modalité orale grâce à diverses stratégies (Maurel *et al.*, 2003). Les deux conclusions de ce travail sont (1) que la transposition vers des patrons intonatifs améliore l'efficacité de la synthèse vocale, mais (2) que l'absence de vision globale et précoce dégrade les capacités cognitives en comparaison avec la lecture silencieuse classique. Une solution alternative ou complémentaire doit être trouvée dans la modalité tactile. Seul quelques travaux de recherches vont dans ce sens.

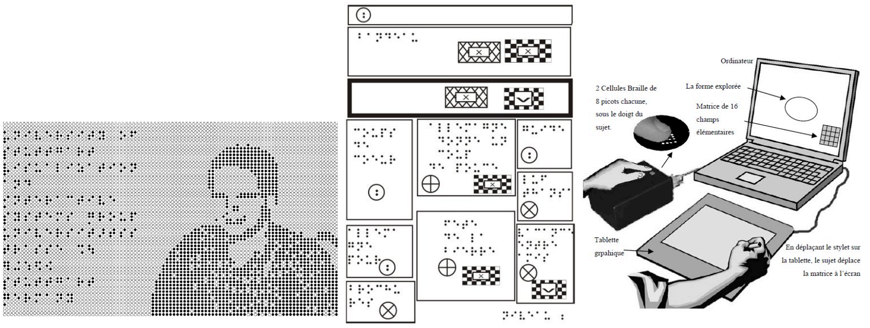


Figure 2. *(a) matrice à picots (b) sortie MAP-RDF (c) Tactos*

D'abord, (Rotard *et al.*, 2005) propose un navigateur web tactile qui affiche les textes et les graphiques bitmaps ou vectoriels sur une matrice de picots afin de restituer la structure bidimensionnelle de l'information (Figure 2a). Bien que cette technique soit intéressante elle nécessite un matériel coûteux et peu utilisable avec des dispositifs portables. De plus, afin de faciliter leur inclusion sociale, les DV préfèrent utiliser les mêmes solutions logicielles ou matérielles que les voyants.

Plus récemment, (Alaeldin *et al.*, 2011) ont présenté un navigateur web tactile à bas coût. Les non-voyants utilisent leur sens tactile pour reconnaître le texte en langage Braille. Une fois de plus, cette solution nécessite un matériel spécifique peu utilisable dans le cadre de l'accès mobile ; de plus cette recherche se consacre à l'accès au texte et non à la perception globale du document.

La solution la plus innovante pour notre problématique (Bou Issa *et al.*, 2009) consiste à détecter l'information visuelle signifiante dans les documents afin d'extraire et réarranger ces propriétés vers une sortie spécifique accessible. Le modèle, MAP-RDF, doit être implanté dans une plateforme logicielle et matérielle capable de fournir des sorties tactiles et orales (Figure 2b). La solution est intéressante puisqu'elle fournit aux non-voyants une vue globale de la structure de la page. Cependant, cette approche ne fonctionne que pour les documents bien structurés qui contiennent des métadonnées et nécessite une capacité d'analyse automatique de la page web qui ne permet pas encore un traitement en temps réel.

Pour résoudre ces difficultés, nous proposons une idée basée sur le travail de (Ziat *et al.*, 2007), dans lequel les auteurs présentent des expériences basées sur le prototype Tactos (Lenay *et al.*, 2003). Cette solution est composée de trois éléments : un ordinateur, des simulateurs tactiles et une tablette graphique avec son stylet (Figure 2c). Chaque forme affichée sur l'écran est perçue de manière haptique selon les mouvements du stylet sur la tablette, guidé par la main de l'utilisateur.

Celui-ci ressent les stimulateurs qui sont activés sous le doigt de l'autre main chaque fois que le curseur traverse une forme sur l'écran.

Nous proposons de la même manière un prototype capable de percevoir de manière haptique la structure visuelle des documents. La mise en page du document doit être analysée (Deng *et al.*, 2003) puis transformée pour que l'utilisateur aveugle puisse balayer l'écran de n'importe quel dispositif (écrans multitouches, pliables, holographiques – Figure 3) avec un capteur qui transforme les contrastes lumineux en vibrations tactiles basse-fréquence.

Figure 3. *(a) écrans pliables (b) écrans holographiques (c) surfaces multitouches*

4 Accès vibrotactile aux documents et aux interfaces

Grâce à un capteur lumineux, l'information du niveau de gris survolé est captée et traduite par un actionneur produisant des vibrations localisées à intensités et fréquences variables. Avec ce dispositif, notre ambition est de remplacer la capacité de balayage visuel de l'écran par une exploration manuelle basée sur la perception vibrotactile. La figure 4, présente notre premier prototype.

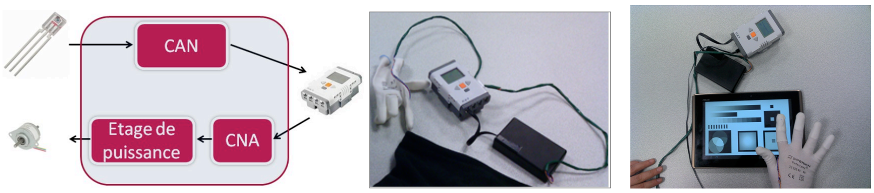


Figure 4. *(a) architecture (b) dispositif vibrotactile (c) dispositif avec la tablette tactile*

Le système réalisé permet de mesurer la lumière sous un doigt en contact avec l'écran et de transformer l'intensité en un stimulus sous la forme d'un niveau de vibration produit par un moteur. Ce moteur, également en contact avec le doigt, permet de « percevoir » l'intensité lumineuse en fonction du niveau de vibration. Actuellement, seule une relation linéaire est utilisée. Le système est conçu pour être simple, économique, aussi flexible que possible et capable de fonctionner sans l'aide de l'ordinateur. L'autonomie du système est supérieure à 24 heures. Un amplificateur opérationnel fournit directement un voltage proportionnel à l'intensité lumineuse reçue par une photodiode intégrée au système dans une boîte opaque qui limite l'incidence de la lumière parasite. Le voltage en sortie du capteur a été numérisé par un convertisseur 8 bits peu coûteux. Enfin, un transistor bipolaire fournit la puissance nécessaire au moteur grâce à des micro-disques de précision. Les fréquences produites sont compatibles avec les capacités physiologiques (avec un maximum autour de 200 Hz). Le coût global du système est inférieur à 400 euros. Un système plus économique est en cours de réalisation en développant un microcontrôleur dédié.

5 Pré-tests d'évaluation du dispositif

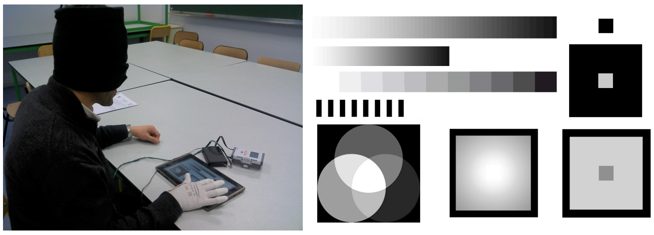


Figure 5. (a) Conditions d'évaluation (b) Conditions d'entrainement

Pour avoir une première idée des possibilités offertes et pour préparer la mise en place d'un véritable protocole d'expérimentation avec des aveugles, nous avons choisi de mettre en place une série d'expériences sur des utilisateurs voyants. 4 tâches ont été proposées et filmées avec 15 sujets voyants les yeux bandés (figure 5a). Chaque sujet a eu une tâche d'entrainement pour « apprivoiser » la perception vibrotactile yeux ouverts sur les formes de la figure 5b. Après l'entrainement suivaient trois tâches d'évaluation. La durée totale de l'expérience était comprise entre 15 et 30 minutes.

5.1 Tâche d'entrainement

L'objectif était de tester le prototype sur divers objets graphiques présentés ensemble sur l'écran. Le sujet pouvait voir et ressentir tactilement différents niveaux de gris, dégradés, pointillés et carrés variant selon leur contour, leur fond, ou leur centre (Figure 5b). Cette tâche permettait d'expérimenter différentes vitesses de balayage sur les carrés en mémorisant les diverses transitions pour accéder à leur centre.

5.2 Tâche de balayage de formes

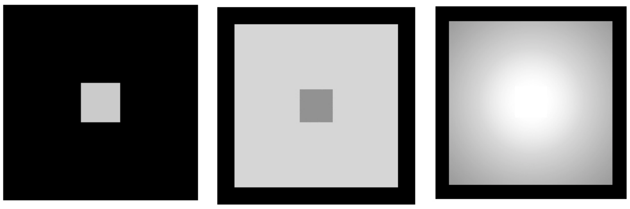


Figure 6. *Formes de base (a) NT2; (b) NT3; (c) NTG;*

Ce second pré-test devait évaluer comment le sujet balaierait certains éléments des formes basiques présentées dans la tâche d'entrainement, mais cette fois en aveugle. Les formes de base de la figure 6 étaient annoncées et présentées successivement à une position aléatoire. Le sujet devait trouver le petit carré au centre du grand. L'intérêt était d'avoir un retour sur l'utilisation des fonds et bordures pour reconnaître suffisamment la forme afin de diriger son doigt. Les 3 formes se différencient par le nombre de transitions qui les compose.

La figure 6a (NT2) nécessite de percevoir deux transitions pour accéder au centre du carré : (1) du blanc au noir et (2) du noir au gris. De la même manière, la figure 6b (NT3) nécessite 3 transitions et la figure 6c nécessite la perception d'un dégradé et donc d'une diminution continue de l'intensité des vibrations du micromoteur. Le temps de balyage était mesuré ; plus d'1 minute était considéré comme une mauvaise réponse. Les tendances empiriques qui sous-tendront nos futures hypothèses sont résumées table 1.

Plusieurs commentaires récurrents méritent d'être clarifiés. Tout d'abord la vibration n'est pas agréable, même si les sujets l'oublient ou la trouve pertinente pour reconnaître les formes. Ensuite, l'exercice nécessite une grande concentration pour ne pas se perdre dans l'espace de l'écran. Enfin, les bordures des formes sont bien perçues.

Ce travail montre que la représentation en dégradée (NTG) est plus efficace pour diriger le doigt vers le centre du carré. La reconnaissance de NT2 est intéressante car l'analyse vidéo montre que le centre est trouvé plus rapidement par une technique de balayage rapide mais que les sujets hésitent à donner la réponse sans avoir raffiné leur recherche. Au contraire, pour NTG, le centre est annoncé dès qu'il est trouvé. Quelques erreurs avec NTG sont à noter en raison de la confusion entre le centre et l'extérieur du carré tous deux blancs. NT3 montre des résultats clairement moins bons.

Forme	**Temps moyen (s)**	**Nombre d'erreurs**	**Ecart-type (s)**
NT2	28	2	19
NT3	36	6	16
NTG	22	3	9

Table 1. *Résultats du pré-test 2*

Les conclusions suivantes peuvent être formulées :
1. un fond noir permet un balayage rapide mais peu confiant ;
2. un fond dégradé permet une reconnaissance confiante mais plus lente ;
3. trop de transitions brutales conduit à une dégradation de la reconnaissance.

5.3 Tâche de reconnaissance de formes

Forme	**Temps moyen (s)**	**Nombre d'erreurs**	**Ecart-type (s)**
NT2	17	1	8
NTG	28	1	16

Table 2. *Résultats du pré-test 3*

Dans cette tâche, deux groupes de 6 sujets devaient trouver et reconnaître en aveugle un des carrés NT2 ou NTG. Le résultat reporté en table 2 montre que les deux formes sont reconnues ; NT2 plus rapidement. On remarque une forte dispersion pour NTG. Nous concluons que :

1. pour reconnaître différentes zones d'un document, des boîtes noires seraient suffisantes ;
2. La sémantique du dégradé peut être utilisée pour une reconnaissance plus fine telle que la distance d'une bordure puisque l'accélération de la vibration est inversement proportionnelle à la longueur du dégradé.

5.4 Tâche de reconnaissance de formes de mises en page inconnues

Notre pré-test final doit permettre de valider l'hypothèse que les pages vibrantes pourraient être envisagées pour accéder à la sémantique des structures visuelles. Le sujet devait reconnaître les formes de la figure 7a, b et c qui ne lui avaient jamais été présentées ; toutes composées de 3 éléments rectangulaires de tailles différentes et positionnés les unes par rapport aux autres de manière particulière. 3 groupes de 5 personnes devaient reconnaître une de ces trois formes. Le temps d'exploration n'était pas limité et a varié entre 2 et 8 minutes. Dès que le sujet estimait avoir suffisamment exploré, il reproduisait sur papier la forme perçue. Le score était évalué sur 9 points :

- 3 points si le nombre d'éléments est respecté (1 point retiré pour chaque forme en défaut ou en excès). La forme en « L » de IDP3 compte pour 2 points ;
- 1 point accordé pour la correspondance entre chaque élément des formes explorées et celles dessinées (maximum 3 points) ;
- 0 à 3 points pour la qualité de la reproduction de la configuration globale (position des éléments les uns par rapport aux autres).

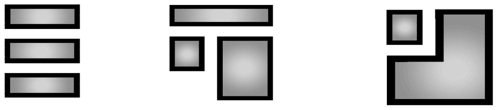


Figure 7. *Formes décrivant des mises en pages (a) IDP1; (b) IDP2; (c) IDP3*

Forme	Score (sur 9)	Ecart-type
IDP1	5.4	3
IDP2	5	2
IDP3	4.8	3

Table 3. *Résultats du pré-test 4*

La diminution graduelle des scores semble correspondre à l'augmentation de la complexité des formes. Les trois sont au-dessus de la moyenne (4.5) mais varient individuellement de 1 à 9. Au moins 1 personne par groupe a obtenu le score maximum de 9 (quelques reproductions en figure 8).

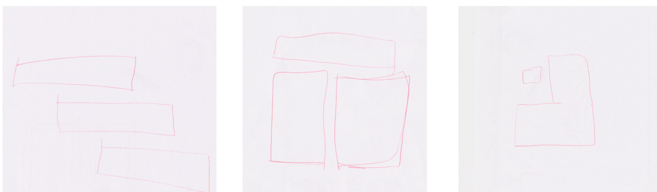


Figure 8. *Formes dessinées (a) pour IDP1; (b) pour IDP2; (c) pour IDP3*

En dépit d'une phase d'apprentissage très courte et une qualité de retour vibrotactile limitée, un voyant, les yeux bandés, peut explorer, identifier et utiliser des modèles pour percevoir la structure visuelle d'un document. Nous souhaitons utiliser ces premiers résultats pour développer des solutions d'accès au « premier regard » d'une page web basé sur un dispositif économique, portable et peu intrusif utilisant le paradigme de la perception vibrotactile.

6 Conclusion et perspectives

Nous avons proposé un nouveau dispositif d'accès non-visuel aux documents électroniques qui :

- produit de nouvelles perceptions basées sur la vibration ;
- est spécifiquement dédié à l'accès de la sémantique des structures visuelles ;
- peut être utilisé sur tout type d'écran ;
- fonctionne en temps réel ;
- est utilisable sur des technologies mobiles.

Les premiers résultats basés sur l'exploration du monde vibrotactile semblent particulièrement intéressants dans le contexte de l'accès précoce aux configurations portées par la typo-disposition des documents. Ce travail est une première étape dans le challenge de fournir aux populations non-voyantes un accès non-linéaire aux textes et aux interfaces. Ce projet, soutenu par l'ANR (ANR-12-SOIN-0003-02) se poursuivra avec un consortium composé d'informaticiens, d'électroniciens et de psychologues pour faire avancer nos connaissances en interfaces homme machine, autour des effecteurs basse-fréquence et sur les capacités cognitives de transposition des modalités sensorielles. Nous espérons parallèlement parvenir à développer une solution économique, portable et non-intrusive qui trouve des perspectives dans le paradigme « design for all ».

Nos premières ambitions en terme d'amélioration du prototype concernent la perception d'un monde plus complexe en multipliant les actionneurs (deux sont nécessaires pour faire correctement du suivi de contour), en en ajoutant d'autres (un actionneur calorimétrique est envisagé pour percevoir les couleurs chaudes *vs.* froides) ; nous pourrons également enrichir la perception en désolidarisant la modulation de fréquence du vibreur de son intensité ; et la rendre plus agréable en introduisant les basses-fréquences adéquates.

Notre ambition est également de passer d'un capteur physique de l'intensité lumineuse à un capteur « logiciel » en introduisant un programme dans la tablette qui donnera les informations sur les pixels survolés au microcontrôleur par Bluetooth ; rendant d'autant moins intrusif le dispositif.

Enfin, nous exploiterons les solutions multimodales en combinant interactions vibrotactiles et perceptions orales ; nos expériences ont d'ailleurs montrées que certains sujets utilisaient aussi bien les vibrations de l'effecteur que le son du moteur pour naviguer dans le monde non-visuel !

Le non-voyant explore le monde en se dirigeant grâce aux contacts de sa canne avec les obstacles et les matériaux autour de lui ; la structure visuelle du texte, par ses variations de gris typographique, peut jouer ce rôle pour l'exploration tactile des documents, en créant un environnement sensoriel fait de « trottoirs textuels », de textures graphiques et de chemins balisés orientant les mouvements de notre « canne-doigt »

7 Références

Alaeldin A., Mustafa Y., Sharief B. (2011). Tactile Web Navigator Device for Blind and Visually Impaired People. In *Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies Conference.*

Bou Issa, Y., Mojahid, M., Oriola, B., Vigouroux, N. (2006). Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents. In *IEEE International Conference on Advances in Computational Tools for Engineering Applications*, 591-594.

Bougie T. (2002). *The Impact of New Technologies on the Quality of Life of People with Disabilities*, ISBN 92-871-5007-9.

Boula de Mareüil, P., Célérier, P., Cesses, T., Fabre, S., Jobin, C., Le Meur, P.-Y., Obadia, D., Soulage, B., Toën, J. (2001). Elan Text-To-Speech : un système multilingue de synthèse de la parole à partir du texte. In *Traitement automatique des langues*, *Synthèse de la parole à partir du texte*, Vol. 42(1), 223-252.

Deng C., Shipeng Y., Ji-Rong W., Wei-Ying M. (2003). Extracting Content Structure for Web Pages based on Visual Representation. In *Proceedings of the the Fifth Asia Pacific Web Conference.*

Earl C., Leventhal, J.,A (1999). Survey of Windows Screen Reader Users: Results And Recommendations. *Journal of Visual Impairment and Blindness*, Vol. 93, num. 3.

Gunderson, J., Mendelson, R. (1997). Usability of World Wide Web Browsers by Persons with Visual Impairments. In *Proceedings of the RESNA Annual Conference.*

Lenay C., Gapenne O., Hanneton S., Marque C. and Genouëlle C. (2003). Sensory Substitution, Limits and Perspectives. In *Touch for Knowing*, John Benjamins Publishers, Amsterdam.

Maurel, F., Vigouroux, N., Raynal, M., Oriola, B. (2003). Contribution of the Transmodality Concept to Improve Web Accessibility. In *Assistive Technology Series Vol.12 (Proceedings of the first International Conference On Smart Homes and Health Telematics – Independent Living for Persons with Disabilities and Elderly People). Chapter 4 - Computer Accessibility*, M. Mokhtari (eds), IOS Press, Amsterdam, 186-193.

Pontelli, E., Gillan, D., Xiong, W., Saad, E., Gupta, G., Karshmer, A. I. (2002). Navigation of HTML Tables, Frames and XML Fragments. In *Proceedings of the fifth international ACM Conference on Assistive Technologies.*

Rotard, M., Knödler, S., Ertl, T. (2005). A Tactile Web Browser for the Visually Disabled. In *Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia*, ACM, New York, NY, USA, 15-22.

Ziat, M., Gapenne, O., Stewart, J., Lenay, C. (2007). Haptic recognition of forms at different scale: A comparison of two methods of interaction. *Interacting with Computers*, Elsevier, Volume 19, Issue 1, 121-132.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 52 (2015) 35 – 42

Procedia Computer Science

The 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015)

Web-Adapted Supervised Segmentation to Improve a New Tactile Vision Sensory Substitution (TVSS) Technology

Waseem Safi, Fabrice Maurel*, Jean-Marc Routoure, Pierre Beust, Gaël Dias

University of Caen Basse-Normandie, GREYC CNRS UMR 6072, Bd Maréchal Juin, 14032 Caen Cedex - France
firstName.lastName@unicaen.fr

Abstract

Nowadays, hand-held devices are being used more and more, especially for web navigation. But the small screen size of these devices requires adapting web page contents to be browsed more conveniently. A fundamental step in a successful automatic adaptation process of a web page is perception its visual layout and mining its Document Object Model (DOM) structure. In this paper, we present a new web-adapted supervised segmentation algorithm dedicated to vibro-tactile access on touch-screen devices. This suggested algorithm is fundamental in our framework whose aim is enhancing the ability of Visually Impaired Persons (VIP) to understand the 2-dimension web page layout by converting web pages into vibrating pages using a graphical vibro-tactile language. A comparison between automatic and manual segmented pages is presented. The objectives of this comparison are, on the one hand, to know how users understand web page layout structure based on their visual perception, and on the other hand, to explore the main differences between automatic and manual segmentation.

Peer-review under responsibility of the Conference Program Chairs

Keywords: *Document Object Model, touch-screen devices, Visually Impaired Persons, vibro-tactile language, layout structure;*

1. Introduction

VIP depend on screen readers in order to deal with computer operating systems and computational programs. One of the most important and desired targets by VIP is navigating the Web, considering the increased importance and

* Corresponding author. Tel.: +33 (0)2 31 56 74 86; fax: +33 (0)2 31 56 73 30.
E-mail address: fabrice.maurel@unicaen.fr

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2015.05.014

expansion of web-based computational programs. Screen readers present some solutions to navigate the textual and graphical contents of web pages, either by transforming a web page into a written Braille, or by transforming the web page into a vocal output. In addition to these solutions, some screen readers installed on touch devices transform a tactile input into a vocal output. There are some drawbacks for these proposed solutions. On the one hand, Braille techniques are costly and only few number of VIP has learned Braille[1], and most of Braille solutions are not suitable for mobile devices[1,2,3]. On the other hand, failure of the proposed solutions to transform the 2-dimension web page structure.

Our work focuses on developing and evaluating a sensory substitution system based on a vibro-tactile solution, which may improve the mentioned drawbacks. In particular, we study how to increase the VIP perception of a 2D web page structure and how to enhance their techniques to navigate the contents of web pages on touch-screen devices. The suggested solution is cheaper than Braille devices and may be efficient in noisy and public environments compared to vocal-tactile solutions. Our contribution is three-fold: (1) designing a Tactile Vision Sensory System (TVSS) represented by an electronic circuit and an Android program in order to transform light contrasts of touch-screen devices into haptic vibrations; (2) designing an algorithm for segmenting web pages in order to support VIP by a way which may enhance their ability to navigate the textual and graphical contents of web pages and (3) exploring how users understand web pages layout structure based on their visual perception.

The paper is organized as follows. First, in section 2, we review rapidly most advanced VIP targeted technologies, and propose our own framework. In section 3, the state of the art for web pages segmentation methods is reviewed. In the fourth section, our hybrid segmentation method is presented. In section 5, we present the results obtained in an experiment where the differences between automatic and manual web page segmentation are compared. Finally, perspectives and conclusions are presented.

2. VIP Targeted Technologies And Proposed Framework

Current products for VIP such as screen readers mainly depend on speech synthesis or Braille solutions, e.g. ChromeVox[3], Window-Eyes[4], or JAWS (Job Access With Speech)[5]. Braille displays are complex and expensive electromechanical devices that connect to a computer and display Braille characters. Speech synthesis engines convert texts into artificial speech, where the text is analyzed and transformed into phonemes. These phonemes are then processed using signal processing techniques. Some screen readers can also support tactile feedback when working on touch-screen devices, such as Mobile Accessibility[6] and Talkback[7] for Android, or VoiceOver[8] for iOS. Many of these products propose shortcuts for blind users to display a menu of HTML elements existing in the web page, for example headers, links and images. But, the main drawback of all these products is the fact that they transfer the information of web pages into a linear way *i.e.* without any indication of the 2-dimension global structure. Many researches tried to enhance the way VIP interact with web pages[9], where a tactile web navigator was proposed. This navigator extracts texts from web pages and sends them to a microcontroller responsible of displaying the text in Braille code using an array of solenoids. A tactile web browser for hypertext documents has been proposed[10]. This browser renders texts and graphics for VIP on a tactile graphics display and supports also a voice output to read textual paragraphs and to provide a vocal feedback.

Another interesting model called MAP-RDF (Model of Architecture of Web Pages) has been proposed[11]. This model allows representing the structure of a web page and provides blind users an overview of the web page layout and the document structure semantics. Finally, Tactos is a perceptual interaction system, which has been suggested [12] and consists of three elements: (1) tactile simulators (two Braille cells with 8 pins) represent a tactile feedback system; (2) a graphics tablet with a stylus represents an input device and (3) a computer. Tactos has been successfully used to recognize simple and complex shapes[13]. However, it shows the same drawbacks (expensive and need specific devices) as the previous systems.

In this work, we aim to increase the ability of VIP to understand the 2-dimension web page layout in order to enhance their tactics to navigate the Web with a vibro-tactile feedback. The first phase in our model is to extract visual structures in the navigated web page and convert these "visual" blocks into zones (or segments) to facilitate the navigation in later phases. We achieve this phase using a hybrid segmentation method. Then the system represents the extracted visual elements as symbols using a graphical language. The third phase is to browse these graphical symbols depending on the size of the used touched-screen device; and in the fourth phase, our system provides a vibro-tactile feedback when the blind user touches the tablet by transforming light contrasts of touch-screen devices into haptic vibrations. To achieve the desired system, we have designed an electronic circuit, which controls four micro-vibrators placed on the fingers. A bluetooth connection with an android tablet allows controlling the vibration frequency and intensity (*i.e.* amplitude) of vibrators. An Android dedicated program on the tablet displays an image on the screen and detects where the user touches the tablet screen (the viewed image represents the result of web page segmentation). The light intensity at touched points is transmitted to the embedded device in order to control the vibration intensity. In this paper, we focus only on the first phase (extracting visual structures in the navigated web page, and convert them into zones). A detailed description of hardware components of the system and results of pre-tests are described in references 2, 14, and 27.

3. Related Works

Segmenting a web page is a fundamental phase for understanding its global structure. Extracting the global structure of web pages is useful in many domains such as information retrieval, data extraction, and similarity of web pages. Many approaches have been suggested for segmenting web pages, such as:

- DOM-based segmentation: it depends on analyzing the DOM tree (Document Object Model), and extracting the main structure of web pages depending on HTML tags[15]. In their paper, the authors determine first the layout template of a web page, and then it divides the page into minimum blocks, and finally collects these minimum blocks into content blocks.
- Vision-based segmentation: this method divides the web page depending on the visual view of web page contents on a web browser. The well known tool VIPS (VIsion based Page Segmentation)[16] is based on this approach.
- Image processing based segmentation: this approach captures an image for the visual view of a web page, and then depends on image processing techniques to divide the captured image into sub-blocks[17,18].
- Text-based Segmentation: this approach focuses on extracting only information about texts in a web page. After dividing the web page into blocks of texts, it could be possible to find the semantic relations between these textual blocks. This method is useful in many information retrieval domains such as question answering applications[19].
- Fixed-length segmentation: this approach divides the web pages into fixed length blocks (passages), after removing all HTML tags, where each passage contains a fixed number of words[20].
- Densitometric analysis based segmentation: this approach depends on methods applied in quantitative linguistics, where text-density refers to a measure for identifying important textual segments in a web page [21].
- Graph-based segmentation: This approach depends on transforming the visual segments of a web page into graph nodes, then applying many common graph methods on these nodes for combining them into blocks, or for making a clustering for these nodes. Some common works which depend on this approach are[22,23].

4. Suggested Hybrid Segmentation Algorithm

Most of segmentation algorithms render first the web page using a web browser, and then segments the HTML elements into many blocks depending on the visual layout. Our constructed hybrid segmentation algorithm has been tested on 154 pages collected manually from many newspapers and e-commerce sites (www.leparisien.fr, www.lefigaro.fr, www.liberation.fr, www.amazon.fr, www.materiel.net, wwww.photobox.fr). The results have been integrated with our under-development Android program. The obtained results are promising because the segmentation algorithm can efficiently extract the web page blocks depending on the visual structure, and the algorithm can also convert correctly these blocks into zones (clustering the blocks). Our algorithm mixes three segmentation approaches, vision-based segmentation, DOM-based segmentation, and graph-based segmentation.

4.1. Vision-Based approach

In this phase, we render the web page using Selenium[24] web driver and Mozilla FireFox browser, and we get its visual structure by Java-script code injection inside the HTML source code of the rendered web page. The obtained visual structure indicates the global hierarchy of the rendered web page. This phase assigns additional information for each DOM HTML element such as XPath and bounding box (location [X0, Y0], and size [height and width]). The input of this phase is a web page HTML source code, and its output is an augmented HTML web page with injected information about bounding boxes and DOM XPath for each HTML element. In next sections, we refer to bounding boxes by blocks (i.e. each bounding box represents an HTML element, and may contain other bounding boxes).

4.2. DOM-Based approach

After obtaining the visual structure of a web page, we analyze its DOM structure by applying filters and re-organization rules for enhancing results of next phases. We divide the DOM elements depending on the specification of HTML5 content models proposed by the World Wide Web Consortium (W3C)[25]. This specification divides the HTML tags into 7 categories (Metadata content, Flow content, Sectioning content, Heading Content, Phrasing content, Embedded content, and Interactive content). The first applied filter is Metadata-Content-Filter, which deletes all the elements considered as metadata content elements except "title" tag. We delete the other tags because they do not contain useful visual information in next steps. The deleted tags are "base", "command", "link", "meta", "noscript", "script", and "style". We do not delete the "title" node (the page title) because the textual information existing in this node could be useful for the user. We add also to this filtered group the following tags "comment", "br", and "doctype".
The second applied filter is Dead-Nodes-Filter, where it deletes all HTML nodes that do not affect on the appearance, for example nodes with height or width equals to "0px" (zero pixel); or nodes with style properties ("display:none" or "visibility:hidden" or "hidden:true"). After applying the previous filters, we apply some re-organization rules in order to enhance visualizing the information in next phases. One example of these rules is Paragraph-Reorganization-Rule, where this rule re-constructs all paragraph child-nodes in one node contains the extracted sub-texts. We made this rule after analyzing many DOM structures, and observing that the text in some paragraph nodes is distributed between many child-nodes such as <i> (italic), <span>, <b> (bold), <strong>, <em> (emphasized), <small>, <mark> (marked), <del> (deleted), <ins> (inserted), <cite> (defining a title of work), <u> (underline), and <sub> (subscript). So extracting these sub-texts and collecting them in one text is useful and more efficient for visualizing them as one block in next phases rather than visualizing them as many separated blocks.
This rule was applied on all tags of type <P> (paragraph), and on all tags existed in the heading content category (specification of HTML5 content models), this group contains the following nodes (h1, h2, h3, h4, h5, h6, and hgroup).

We also apply this rule on many other HTML tags which might contain textual child-nodes such as <a> (hyperlink tag), <abbr> (abbreviation), <acronym> (this tag is not supported in HTML5), <address> (contact information for the author/owner), <bdi> (Bi-directional Isolation), <button>, <label>, <li> (list element), and <q> (quotation). We used Jsoup tool[26] in order to access to a web page DOM structure, and getting its HTML hierarchy. The result of this phase is a filtered DOM-tree; each of its nodes is visible and contains XPath and bounding box information. The designed filters and re-organization rules were integrated with our framework, and then we applied these rules and filters on the vision-based segmented web pages (154 pages mentioned previously). After getting the filtered DOM-tree for each page, we represented the obtained bounding boxes on the used tablet Samsung GALAXY Tab 2 (10.1inch, dimensions HeightXWidthXDepth 175.3X256.7X9.7 mm, Android version 4.1.2), after making a scaling of sizes of bounding boxes to be appropriate with the new size of used tablet.

4.3. Graph-Based approach

After segmenting the web page depending on its visual structures and analyzing its DOM-structure, we apply a new graph-based segmentation algorithm called "Blocks2Zones Clustering" in order to group many similar blocks together in one zone. Clustering many blocks together is necessary in order to decrease the number of viewed blocks in some interfaces (instead of viewing many blocks, we view one zone which represents these blocks and then the user can navigate intra-elements inside the zone by double clicking on the graphical element of the chosen zone), and to group closed blocks in one zone (here, closeness depends on distances between blocks, this will be described next sections in details). The pseudo-code of the proposed algorithm is:

Blocks2Zones Clustering Algorithm
Input (Blocks, N° of desired Zones)
Output: Graph of N nodes (N Zones)
1- Transform the blocks into a graph (Un-Directed graph)
1.1. Blocks → Nodes,
1.2. Make relations between the nodes, and assign weights for these relations.
2- If number of zones <= number of blocks
end the algorithm,
Else
3- Find the node with the smallest size (node A)
(size of node ==size of the rectangle bounds the node)
4- For node A, find the connection which has the largest weight (node B).
5- Group the nodes A, and B (A+B).
6- Repeat steps 3-4-5 till number of blocks == number of zones

The output graph is described as following: G = (V, E), where G is undirected graph, V is a set of vertices (nodes or zones), and E is a set of edges (connections between zones), and |V|=n (number of desired zones). We define the set of vertices as V={$v_{i:}$ 1≤i≤n where n is the number of desired zones, and v_i is set of sub-zones}. We define the set of connections as following E={e_j: $e_j(v_{j1},v_{j2})$: v_{j1}∈V, and v_{j2}∈V}. To calculate weights between nodes, we used the Euclidian distance ($d(p,q) = d(q,p) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$). Where p and q are centers of two nodes between them we want to calculate the distance (center of a node is the center of the rectangle which bounds the node).

In following, an example of applying this algorithm on the main page of web site w3schools.com is given. In this example, we want that the algorithm segments the main page into 9 zones. The first iteration of the algorithm will divide the page into 13 zones (because there are 13 main nodes in the DOM structure – <Div> nodes) as illustrated in figure 1.a, and will construct a graph of 13 nodes as illustrated in figure 1.b.

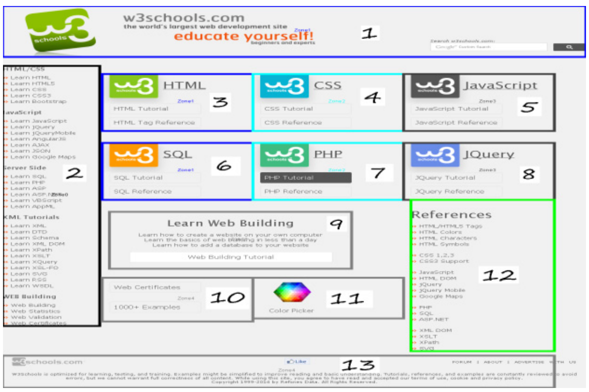


Fig. 1. (a) Segmentation for 13 zones of www.w3schools.com

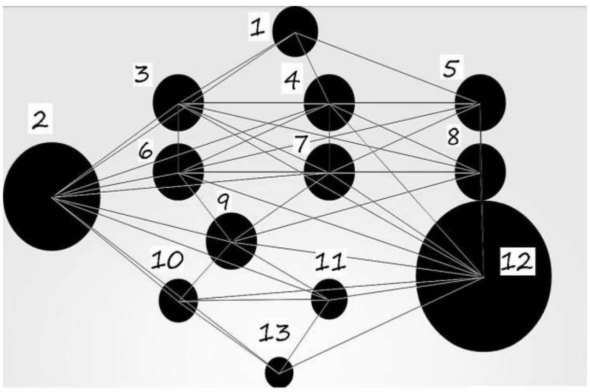


Fig. 1. (b) Constructed graph

Figures 2 and 3 illustrate next iterations of the algorithm and how the 13 zones are converted to 9 zones.

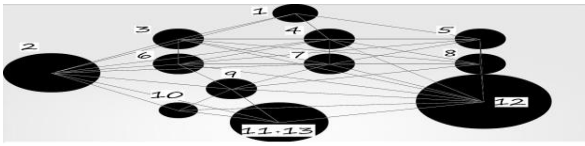


Fig. 2. (a) 12 zones segmentation

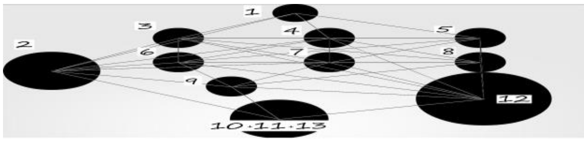


Fig. 2. (b) 11 zones segmentation

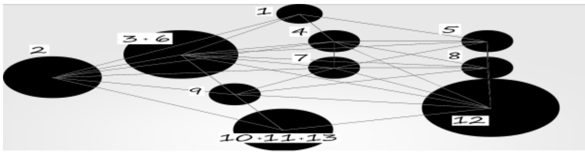


Fig. 3. (a) 10 zones segmentation

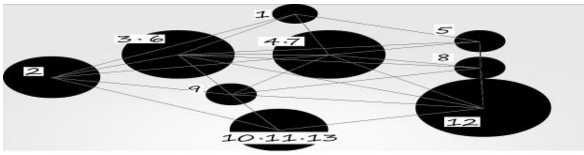


Fig. 3. (b) 9 zones segmentation

Applying this hybrid segmentation algorithm on a filtered DOM-tree (obtained from applying Vision-based approach and DOM-based approach) converts a web page to a set of zones, each zone contains many other zones or blocks, since each block represents a visual structure of HTML element and may contain many other blocks. The purpose of the proposed vibro-tactile access protocol is then to transform semantics of symbols in these zones, or blocks, or HTML elements into vibrations with different frequencies and amplitudes.

5. Manual And Automatic Web Page Segmentation Differences

In order to evaluate our algorithm results, and to know how users understand web layout structures based on their visual perception, we made an experiment as following: we asked to 15 volunteers a manual segmentation of different kinds of web pages. The volunteers were of different ages (between 25 and 50 years old), and most of them were informatics specialists. We presented, for each volunteer, 4 printed copies (A4 size papers) of 8 web pages (2 pages from www.cdiscount.com, 2 pages from www.photobox.com, 2 pages from www.rueducommerce.fr, 1 page from www.w3schools.com, and 1 page from www.leparisien.fr). We asked each volunteer to segment the 4 copies of each web page into 3, 4, 5 and 6 zones with following considerations:

- all the pages are printed in gray scale in order to avoid affecting the colors on the segmentation process. We chose this option because the current version of our algorithm does not depend on color differences between blocks,
- users can segment the page using polygons with minimum of 4 points (triangles are not allowed),
- users can start segmenting from any part or direction of the page (left, right, top, down, or center),
- we asked users to write the ordering number of zones (inside or beside the zone) while they make the segmentation process; this is very useful for us to know how the users start segmenting the pages, and how they end it.

After collecting all manual segmented copies (480 papers : 15 users X 8 web pages X 4 copies), we noticed the following:

- 70% of users do not start the segmentation process for certain number of zones by the same way they start segmenting for other numbers of zones,
- 16% of papers have been segmented starting from the center of the page, 3.5% have been segmented starting from the bottom of the page, and the majority of papers 80.5% have been segmented starting from top of the pages,
- 20% of papers have been segmented vertically, and 80 have been segmented horizontally,
- 92.6% of segments are rectangles, and 7.4% of segments are polygons with more than 4 points,
- finally, we noticed that it is very difficult to detect a segmentation method common between all users; since each user segments the pages depending on his understanding of the web page layout structure, on his visual perception of the visible elements, and on his interests and visual experience.

We have run our algorithm with the 8 mentioned web pages. The algorithm segmented each web page into 3, 4, 5, and 6 zones. The comparison between manual and automatic segmentation is illustrated in tables 1 and 2.

Table 1. Strong-criterion based matching results of manual and automatic segmentation

	Page1	Page2	Page3	Page4	Page5	Page6	Page7	Page8	Total
3 Zones	1	0	0	6	3	0	9	1	20
4 Zones	1	5	4	3	3	0	3	2	**21**
5 Zones	4	4	1	4	3	0	0	2	18
6 Zones	4	1	2	2	4	1	0	1	15
Total	10	10	7	15	13	1	12	6	74

Table 1 illustrates matching results based on what we call “Strong criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if the results are 100% identical without any difference. We can conclude from table 1 that pages segmented automatically and manually into 4 zones are more matched than pages segmented into other numbers of zones. The percentage of identical matching depending on the strong criterion is 15.42% (74 identical matched results of 480 segmented copies). Table 2 illustrates matching results based on what we call “Weak criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if at least 50% of the results are identical. We can conclude from table 2 that pages segmented automatically and manually into 3 zones are more matched than pages segmented into other numbers of zones. The matching percentage depending on the weak criterion is 47.5% (228 weak identical matched results of 480 segmented copies).

Table 2. Weak-criteria based matching results of manual and automatic segmentation

	Page1	Page2	Page3	Page4	Page5	Page6	Page7	Page8	Total
3 Zones	8	10	8	9	9	2	12	6	**64**
4 Zones	7	5	7	8	10	4	13	7	61
5 Zones	8	6	6	5	7	6	8	9	55
6 Zones	6	4	7	5	8	5	10	3	48
Total	29	25	28	27	34	17	43	25	228

Figures 4.a and 4.b illustrate matching results as percentages of the manual segmented pages, based on the “Strong criterion” and the “Weak criterion” (the horizontal axe indicates the segmented pages and the number of zones, and the vertical axe indicates percentage of matching for automatic segmented pages with manual segmented pages).

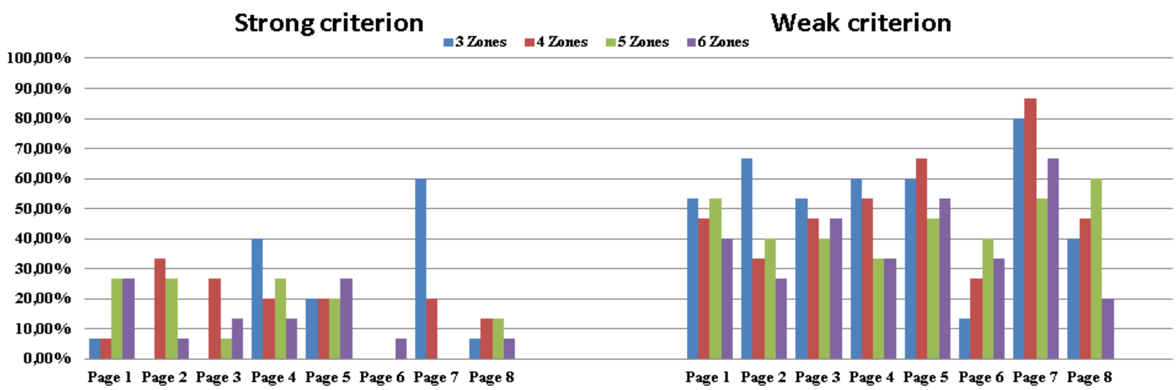


Fig. 4. (a) Percentage of Weak-criteria based matching results

Fig. 4. (a) Percentage of Strong-criteria based matching results

Figures 4.a and 4.b represent percentage of matching (of each page for each zone). The percentage has been calculated via dividing the number of matching by the number of users (15). For example, in figure 4.a, for the page number 7, the percentage of matching the automatic 3-zones segmentation with the manual segmentation is 60% (9 matching /15 users). These results are promising considering that this is the first version of the algorithm, and especially that this version does not take into account the human visual perception aspects, and does not use image processing techniques which could be very useful in enhancing the results.

6. Conclusion and Perspectives

In this paper, we presented our work whose aim is enhancing the ability of VIP to understand the 2-dimension web page layout, and we focused on the proposed supervised segmentation algorithm. Next steps in this research will be 1) adding some image processing techniques in order to enhance the proposed segmentation algorithm 2) adding advanced techniques in text summarization to facilitate navigating textual information, 3) making real experiments with blind VIP to study effects of suggested segmentation algorithm on their web navigation models.

References

1. F. Maurel, G. Dias, J-M. Routoure, M. Vautier, P. Beust, M. Molina, and C. Sann, "Haptic Perception of Document Structure for Visually Impaired People on Handled Devices", Procedia Computer Science, Volume 14, Pages 319-329, ISSN : 1877-0509, 2012.
2. F. Maurel, N. Vigouroux, M. Raynal, and B. Oriola, "Contribution of the Transmodality Concept to Improve Web Accessibility". In Assistive Technology Research Series, Volume 12, 2003, Pages 186-193. International conference; 1st, Smart homes and health telematics; Independent living for persons with disabilities and elderly people. ISSN : 1383-813X. 2003.
3. http://www.chromevox.com/ [Access 10/12/2014]
4. http://www.synapseadaptive.com/gw/wineyes.htm [Access 10/12/2014]
5. http://www.freedomscientific.com/ [Access 10/12/2014]
6. https://play.google.com/store/apps/details?id=es.codefactory.android.app.ma.vocalizerfrfdemo&hl=fr [Access 10/12/2014]
7. https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=fr [Access 10/12/2014]
8. http://www.apple.com/fr/accessibility/ [Access 10/12/2014]
9. A. Alaeldin, Y. Mustafa, and B. Sharief, 2012. "Tactile WebNavigator Device for Blind and Visually Impaired People". In Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies, Jordan, 2011.
10. M. Rotard, S. Knödler, and T. Ertl, "A Tactile Web Browser for the Visually Disabled". In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, 2005, pages 15-22, 2005.
11. Y. BouIssa, M. Mojahid, B. Oriola, and N. Vigouroux, "Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents". IEEE International Conference on Advances in Computational Tools for Engineering Applications, Pages: 591 – 594, 2009.
12. C. Lenay, O. Gapenne, S. Hanneton, C. Marque, and C. Genouëlle, "Sensory Substitution, Limits and Perspectives". In Touch for Knowing Cognitive psychology of haptic manual perception, Amsterdam, Pages: 275-292, 2003.
13. M. Tixier, C. Lenay, G. Le-Bihan, O. Gapenne, and D. Aubert, "Designing Interactive Content with Blind Users for a Perceptual Supplementation System", TEI 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain, Pages 229-236, 2013.
14. F. Maurel, W. Safi, P. Beust, and J.M. Routoure, "Navigation aveugle sur dispositifs mobiles : toucher le Web... pour mieux l'entendre", 16ème Colloque International sur le Document Électronique, CIDE16, Lille, France, Europia productions, 2013.
15. A. Sanoja, and S. Gançarski, "Block-o-Matic: a Web Page Segmentation Tool", BDA. Nantes, France. 2013.
16. C. Deng, Y. Shipeng, W. Ji-Rong, and M. Wei-Ying, "VIPS: a Vision-based Page Segmentation Algorithm", Nov. 1, 2003, Technical Report MSR-TR-2003-79, Microsoft Research. 2003.
17. D. Cai, X. He, W-Y. Ma, J-R. Wen, and H. Zhang, "Organizing WWW Images Based on the Analysis Of Page Layout And Web Link Structure", Microsoft Research Asia, Beijing, China, 2004.
18. J. Cao, B. Mao, and J. Luo, "A segmentation method for web page analysis using shrinking and dividing", International Journal of Parallel, Emergent and Distributed Systems - Network and parallel computing, Volume 25 Issue 2, April 2010. Pages: 93-104, 2010.
19. N. Foucault, S. Rosset, and G. Adda, "Pré-segmentation de pages web et sélection de documents pertinent en Questions-Réponses", TALN-RÉCITAL 2013.
20. J.P. Callan, "Passage- level Evidence in Document Retrieval", the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Pages: 302-310. Publisher: ISBN:0-387-19889-X, Springer-Verlag New York, Inc, 1994.
21. C. Kohlschütter, and W. Nejdl, "A Densitometric Approach to Web Page Segmentation", USA, CIKM'08, Proceedings of the 17th ACM conference on Information and knowledge management, Pages: 1173-1182. 2008.
22. D. Chakrabarti, R. Kumar, and K. Punera, "A graph-theoretic approach to webpage segmentation". Proceedings of the 17th international conference on World Wide Web, WWW'08, ACM, USA, Pages: 377-386. Publisher: ACM New York, NY, USA, 2008.
23. X. Liu, H. Lin, and Y. Tian, "Segmenting Webpage with Gomory-Hu Tree Based Clustering", International Journal of Software Engineering and Its Applications, Vol 6, No 12, Pages: 2421-2425. 2011.
24. http://www.seleniumhq.org/ [Access 10/12/2014]
25. http://www.w3.org/TR/2011/WD-html5-20110525/content-models.html [Access 10/12/2014]
26. http://jsoup.org/ [Access 10/12/2014]
27. W. Safi, F. Maurel, J-M. Routoure, P. Beust, and G. Dias, "Blind Browsing on Hand-Held Devices: Touching the Web to Understand it Better", Data Visualization Workshop associated to 25th ACM Conference on Hypertext and Social Media (HYPERTEXT 2014), 2014.

Blind Navigation of Web Pages on Touch-Screen Devices

Waseem SafI Fabrice Maurel Jean-Marc Routoure Pierre Beust Gaël Dias
University of Caen Basse-Normandie
GREYC CNRS UMR 6072
Bd Maréchal Juin
14032 Caen Cedex - France
firstName.lastName@unicaen.fr

Abstract. Nowadays, Touch-Screen devices are being used more and more, especially for web navigation. But the small screen size of these devices requires adapting web page contents to be browsed more conveniently. A fundamental step in a successful automatic adaptation process of a web page is perception its visual layout and mining its Document Object Model (DOM) structure. Considering that navigating the Web is one of important missions in the field of computer accessibility. Many specialized techniques for VIP (Visually Impaired People) succeeded to extract the information displayed on digital screens and succeeded to transform this information in a linear way either into a written format on special Braille devices, or into a vocal output using text to speech synthesizers. However, although this success, screen readers failed to transform the 2-dimentional structure of the navigated web page; despite many researches confirm that perception the structure enhances web navigation and memorization. In this paper, we propose a new technique aimed to enhance the VIP ability to navigate the Web by affording a "first glance" web page overview. This technique focuses on improving non-visual vibrotactile access to web pages on touch-screen devices, based on extraction and re-organization the structure of texts and graphical elements for web pages, reformatting and converting automatically these visual structures and textual information into vibrating pages using a graphical vibro-tactile language. We present also a new web-adapted supervised segmentation algorithm dedicated to vibro-tactile access on touch-screen devices. This suggested algorithm is fundamental in our framework. A comparison between automatic and manual segmented pages is presented. The objectives of this comparison are, on the one hand, to know how users understand web page layout structure based on their visual perception, and on the other hand, to explore the main differences between automatic and manual segmentation.

Keywords: Touch-Screen devices, Document Object Model, Visually Impaired People, first glance, vibro-tactile.

1 Introduction

In October 2013, the world health organization estimated that the number of VIP in the world is 285 million, 39 million of them are blind, and 246 million of them have low vision. The organization defined four levels of visual functions depending on the international classification of diseases, which are: normal vision, moderate visual impairment, severe visual impairment, and blindness*. VIP depend on screen readers in order to deal with computer operating systems and computational programs. One of most important and desired targets by VIP is navigating the Web, considering the increased importance and expansion of web-based computational programs. Screen readers present some solutions to navigate the Web, either by transforming a web page into a written Braille, or into a vocal output. Some screen readers installed on touch devices transform a web page into a Vocal-Tactile output. But there are some drawbacks for these proposed solutions: on the one hand, the Braille techniques are costly, and only few number of VIP have learned Braille (in France, there are about 77 000 visually impaired people and only 15 000 of them have learned Braille -from statistics published in September 2011-)**. On the other hand, transforming the information of a web page into a vocal format might not be suitable in public and noisy environments. Finally most of Braille solutions are not suitable for mobile devices [1]. In addition to these drawbacks, the most important one is the failure to transform the 2-D web page structure, because as reported by many authors, perception the 2D structure greatly improves navigation efficiency and memorizing the information because it allows high level reading strategies (rapid or cursory reading, finding or locating information,...) [2]. Our work focuses on developing and evaluating a sensory substitution system based on vibro-tactile solution which may solve the mentioned drawbacks; where we study how to increase the VIP perception of a 2-D web page structure, and how to enhance their techniques to navigate the Web on touch-screen devices. This suggested solution is very cheap comparing with prices of Braille devices, and also it could be more efficient in noisy and public

*http://www.who.int **http://www.opc.asso.fr/

environments comparing with vocal-tactile solutions. Our contribution is three-fold:
- Designing a Tactile Vision Sensory System (TVSS) represented by an electronic circuit and an android program in order to transform light contrasts of touch-screen devices into low frequencies tactile vibrations, and running a series of experiments with blind persons in order to validate our hypotheses,
- Designing an algorithm for segmenting web pages automatically in order to support the VIP by a way which may enhance their ability to navigate the textual and graphical contents of web pages, and
- Running a series of experiments with sighted persons in order to compare the differences between automatic and manual web page segmentation.

The paper is organized as following: first, in section 2 we view the state of the art for VIP targeted technologies, and we propose our own framework. The first pre-tests achieved with blind persons will then be described in the third section, and an analysis of results will be presented in the fourth section. In section 5, the state of the art for web pages segmentation methods is reviewed. In section 6, our supervised hybrid segmentation method is presented. In section 7, we present the results obtained in an experiment, where the differences between automatic and manual web page segmentation are compared. Finally, perspectives and conclusions are presented.

2 VIP Targeted Technologies and Proposed Framework

Current products for VIP such as screen readers depend mainly on speech synthesis or Braille solutions, such as ChromeVox[3], Windows-Eyes[4], and Jaws (Job Access With Speech)[5]. Braille displays are complex and expensive electromechanical devices that connect to a computer and display Braille characters. Speech synthesis engines convert texts into artificial speech, where the text is analyzed and transformed into phonemes; these phonemes are then processed using signal processing techniques. Some screen readers can support a tactile feedback when working on touch devices, such as Mobile Accessibility[6], Talkback[7] for Android, and VoiceOver[8] for IPad. Many of these products propose shortcuts for the blind user to display a menu of HTML elements existed in the web page, for example headers, links, and images. But, the main drawback of all these products is that they transfer the web page information into a linear way, and without any indication for the global web page structure (2D layouts). Many researches tried to enhance the way by which VIP interact with web pages, such as [9], that proposed a tactile web navigator to enable blind people to access the Internet. This navigator extracts texts from web pages, and sends these texts to a microcontroller responsible of displaying the text in Braille language using an array of solenoids.

A tactile web browser for hypertext documents has been proposed by [10]. This browser renders texts and graphics for VIP on a tactile graphics display, and it supports a voice output to read textual paragraphs and to provide a vocal feedback. The authors implemented two exploration modes, one for bitmap graphics, and another one for Scalable Vector Graphics. Main drawback of this proposed system is that it needs a pin matrix device, which is expensive and cannot be integrated with handled devices. Another interesting model called MAP-RDF ("Model of Architecture of web Pages") [11] proposed a method to improve the accessibility to visual information for blind persons. This model allows representing the structure of a web page, and provides the blind users with an overview of the web page layout and the document structure semantics. The main drawback of this model is that it could be applied only on well structured web pages which contain meta-data, so it could not be applied to most web pages which rarely contain meta-data.

Tactos is a suggested perceptual interaction system [12]. It consists of three elements: 1- tactile simulators (two Braille cells with 8 pins) represent a tactile feedback system, 2- a graphics tablet with a stylus (represents an input device), 3- computer [13], The graphics tablet and the stylus allow the user to explore graphical contents on the screen such as circles, rectangles, and characters. While the user explores the contents, the system transforms pixels under the stylus into tactile stimulation on the Braille cells. 30 prototypes of Tactos have been released, to be used by a lot of users in many domains. Tactos has been successfully used to recognize simple and complex shapes. The device has been also used in geometry teaching domain in an institution for visually impaired and blind children. Tactos also allowed psychology researchers to propose and develop new paradigms for studying perceptions and mediated communication of blind persons [13].

First glance could be defined as the ability -in a blink of an eye- to understand the document layout and its structural semantics [1]. We aim of our work to increase the ability of visually impaired persons to understand the web page 2-dimentional layout in order to enhance their tactics to navigate the Web. A commercial tablet connected to a vibro tactile set-up is used for that. The first phase in our model is to extract visual structures in the navigated web page, and to convert these visual blocks to zones (segments) for facilitating the navigation in later phases. We achieve this phase depending on a hybrid segmentation method. Then the system will represent on the tablet screen the extracted visual elements as symbols using a graphical language (this language is under-development). The third phase is to browse these graphical symbols depending on size of the used touched screen device, and then in the fourth phase, our system provides a vibro-tactile feedback when the blind user touches the tablet. The intensity and the frequency of the vibration depend mainly on gray level under the finger.

In this paper, we focus only on the first and fourth phases which specialize in extracting visual structures in the navigated web page, and in giving the user a vibro-tactile feedback by transforming light contrasts of touch-screen devices into low-frequencies tactile vibrations. To achieve the desired system, we have designed an electronic circuit which controls two micro-vibrators placed on the hands. A Bluetooth connection with an android tablet allows controlling the vibration intensity (Amplitude) of vibrators. An Android dedicated program on the tablet displays an image on the screen and detects where the user touches the tablet screen. The gray level of touched points is transmitted to the embedded device in order to control the vibration intensity. Only one micro-vibrator was used for pre-tests described in this paper. Figure 1 illustrates the designed electronic circuit, and the used vibrator.

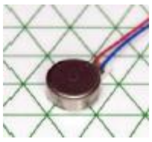


Figure 1. (a) The used micro-vibrator

Figure 1. (b) The designed electronic circuit

Figure 1. (a) The used micro-vibrator: the range of vibration frequency goes from 20 Hz up to 260 Hz. (b). The embedded system designed with the Bluetooth module.

3 Pre-tests Protocol

3.1 Objectives of Pre-tests

Our objective of the designed protocol is enhancing the ability of VIP to recognize the 2-D structure of a web page. In order to test the prototype mentioned in section 2, we designed some images contain different structures (detailed in section 3.2), and we tested the prototype firstly on 15 sighted persons (their eyes were closed) [13], and later on 5 blind persons. Testing the protocol on sighted and blind persons gave us a more understanding of tactics and strategies achieved by sighted and blind persons to navigate the designed structures. This will be useful in designing the desired graphical vibro-tactile language (all results are detailed in next sections).

3.2 Designed Protocol for Vibro-Tactile Access

Each experiment (either for sighted or blind persons) consists of 4 ordered phases of training (learning task), and four ordered phases of evaluation (evaluation task). All the experiments were filmed, and the designed program stocked many parameters in log files (coordinates X, Y, pressure on the screen, and the time at each touch). Figure 2 presents the 4 images of training phases, and figure 3 presents the 4 images of evaluation phases.

Figure 2 (a). Image a of training task

Figure 2 (b). Images b (NT2), c (NT3), d (NTG)

Figure 2. Images of training task

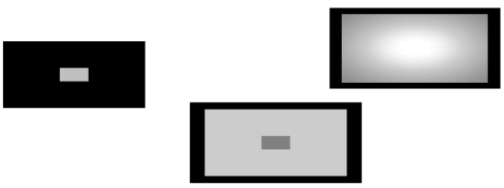


Figure 3 (a). Image a (Evaluation task)

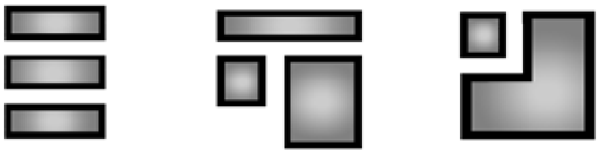


Figure 3 (b). Images b (IDP1), c (IDP2), d (IDP3)

Figure 3. Images of evaluation tasks.

In the training task, each user discovered firstly the graphical elements in each image presented in figure 2 (images a, b (NT2), c (NT3), and d (NTG)), and users were informed about names of graphical elements. The name of each image NT2, NT3, NTG, indicates how many transitions are necessary to access the square center, for example NT2 proposes 2 transitions to access the center of the square.
The evaluation task consists also of 4 phases, the first one allows to discover the image 3.a and to name each square inside it, then next phases are about images 3.b, 3.c, and 3.d, where we asked users to discover contents of each image, then to describe these contents, and to redraw discovered elements inside each image. We chose these images depending on following considerations:
- Image 3.a contains all squares on which users have trained in the training task, so it could test the ability to memorize and to distinguish the shapes.
- Image 3.b contains 3 rectangles with matched sizes and with vertical order, and the image 3.c contains 3 rectangles with different sizes and many relations of directions, so testing images 3.b, and 3.c could test the ability of distinguishing sizes, and distinguishing relations of directions.
- Image 3.d contains different shapes (a rectangle and a polygon), so it could test the ability to distinguish different shapes in the same image.
- The tested images contain examples of expected results of the segmentation process, so success of distinguishing these shapes by blind users could be an indicator of their ability to distinguish results of segmenting web pages.
The results of pretests with sighted persons were already published in [1]. Table 1 presents some results of the experiment for images NT2, NT3, NTG (the time required to distinguish graphical elements and the number of errors for the 15 sighted persons. The users have been asked to name the shapes in figure 3.a, and for each shape, we evaluated the number of correct and incorrect answers). In table 1, number or errors represents the number incorrect answers.

Table 1. Results of experiments of sighted persons for images NT2, NT3, NTG. (Table extracted from [1])

Shape Name	Average Time in Seconds	Number of errors
NT2	28	2
NT3	36	6
NTG	22	3

We notice from table 1 that the lowest number of errors is assigned to image NT2, and the largest time and max number of errors is assigned to image NT3.

3.3 Experiment Steps with Blind Persons

The tests performed with each one of the 5 blind persons consisted of following: personal and technical questions, explanations of the test objective, a training task, and finally an evaluation task. The approximated average time for the test for each person is about 1 hour.

3.3.1 Personal and technical questions

Before starting the tests with the 5 blind persons, we asked them to support us with some information about their age and date of their blindness. Table 2 summarizes answers of personal questions.

Table 2. Personal information of the blind persons

User-ID	0	1	2	3	4
Age (Years)	63	67	59	56	36
Sex	Male	Female	Male	Female	Female
Date of the blindness	since birth	since 32 years	since 25 years	since 10 years	since 15 years

We also asked users to provide us with some technical information about their experience in dealing with operating systems, screen readers, and what are the main problems when they navigate the Web. Table 3 shows a summary of answers for these technical questions. The first two columns indicate the number of operating systems (either Windows or Linux) used either on fixed or portable computers. The third, fourth, and fifth columns indicate the number of users who use JAWS (Job Access With Speech), NVDA (NonVisual Desktop Access), and ORCA, either on fixed or portable computers.

Table 3. Used operating systems and screen readers

	Windows	Linux	JAWS	NVDA	ORCA
Number of users with fixed computer	4	0	4	2	0
Number of users with portable computer	2	1	2	1	1

No one of the five blind persons uses a tablet, and the screen readers used with cellular phones are Talks and MobileSpeak with Nokia, and Voiceover with IPhone. Only one of the 5 persons uses a telephone to access the Web (access via IPhone). The main problems of accessing the Web via fixed or portable computers, or via IPhone telephone were: problems of access to Flash files, problems of AJAX technologies, and no ability to know the global structure of web sites. (These problems have been reported to us by the 5 users).

3.3.2 Explaining the objective of the test

To give the blind persons a good idea about the test, we explained in details what are the objectives and the phases of each tasks, and described contents of the embedded system; we also explained the final objectives of the project, and why we concentrate on vibro-tactile technique regardless of other techniques. This phase was important to initiate users for accepting kindly the test and for doing their best to interact with next steps as correctly as possible.

3.3.3 Training and learning task

In this training task, the user discovers the graphical elements in each image presented in figure 4 (images a, b (NT2), c (NT3), and d (NTG)), and the users were informed of each shape name. This task was very important for users to test the system before the evaluation task, and to know exactly how the system transforms different the grey level under the touched points on the tablet screen to a vibration mode. It is also very useful for users to control their speed of mapping the screen either to discover either the borders or the contents. During this task, the program recorded the touching information in log files (X, Y coordinates, Pressure, and Time). Table 4 indicates training times in minutes for each user, and for each image in figure2.

Table 4. Times of training task for each user (in minutes)

User ID / Image	ID0	ID1	ID2	ID3	ID4	Total	Average
a	4,99	2,60	3,55	9,97	3,96	25,08	5,02
b (NT2)	4,40	3,11	0,99	3,30	1,00	12,80	2,56
c (NT3)	2,81	6,29	2,54	2,85	1,15	15,65	3,13
d (NTG)	3,11	3,98	2,02	3,22	1,32	13,65	2,73
Total	15,31	15,99	9,10	19,34	7,43		

We notice from table 4 that discovering the first image takes more time, and it is normal because it is the first experiment for blind users on this prototype. We notice also that there is a significant decrease in time between discovering the first and the last image in the training task. This could be an indicator that training users could decrease the time for discovering graphical elements. We can also notice the significant difference between different tested persons, for example user with ID4 needed 7.43 minutes to scan the images (A, NT2, NT3, and NTG), but the user with ID3 needed 19.34 minutes to scan the same images.

3.3.4 Evaluation task

In this task, firstly we asked each user to discover the image 3.a and to find how many squares inside it and to name each founded square, then we asked them to discover images 3.b (IDP1), 3.c (IDP2), 3.d (IDP3), and to describe them to us, and to redraw discovered shapes. Table 5 illustrates an evaluation of answers for the first question to name squares in image 3.a (The blind users have been asked to name the shapes in image 3.a, and for each shape, we evaluated the number of correct and incorrect answers). (In tables 5 and 6, the symbol ✓ represents a correct answer for the touched shape, and the symbol X represents an incorrect answer or inability to select the name of the touched shape).

Table 5. Results of questions for squares in image 3.a

User-ID/ Square Name	ID0	ID1	ID2	ID3	ID4	Number of errors
NT2	✓	✓	✓	X	✓	1
NT3	✓	✓	X	X	✓	2
NTG	✓	X	✓	X	✓	2

We notice from table 5 that the lowest number of errors is assigned to image NT2, and it is the same result which we obtained during tests with sighted persons. Results of answers for other questions related to images IDP1, IDP2, and IDP3 are summarized in table 6.

Table 6. Answers of questions for images IDP1, IDP2, IDP3

User-ID	ID0			ID1			ID2			ID3			ID4		
IDP	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Answers about number of rectangles	✓	✓	X	✓	X	X	✓	X	✓	X	✓	✓	✓	✓	✓
Answers about sizes of rectangles	X	X	X	X	X	X	✓	X	X	X	✓	✓	✓	✓	✓

We notice from data in tables 4, 5, and 6, that the best performance is for the user with ID4, and this may be because that this female user is the youngest between others, and it could be because that she was the only one that has already used touched devices (an IPhone in her case working with VoiceOver). After answering questions about each image of images (IDP1, IDP2, IDP3), we asked each user to redraw the graphical elements founded in each touched image. Figure 4 views the redrawing results of the user ID4 (ID4 is the female user who gave best answers).

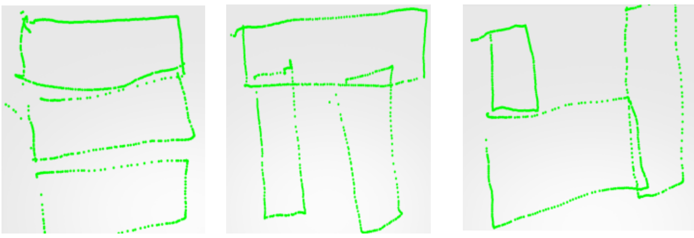


Figure 4. Results of redrawing images IDP1, IDP2, IDP3 for the user ID4

When comparing results of redrawing (Figure 4) with images IDP1, IDP2, and IDP3, we find that the results are interesting, and we can conclude the following:

1. An ability of distinguishing sizes of shapes, because the degree of scaling between redrawn shapes is nearly equal to the degree of scaling between real shapes (IDP1, IDP2, IDP3).
2. An ability of distinguishing relations of directions, because relations of directions (vertical order, left to, right to,…) between redrawn shapes is nearly equal to relations of directions between real shapes.

The average of times in minutes consumed for each evaluation question is summarized in table 7.

Table 7. Times of the evaluation task for each user (in minutes)

User ID / Image	ID0	ID1	ID2	ID3	ID4	Total	Average
A	1,23	9,87	5,39	7,84	1,85	26,17	5,23
IDP1	4,39	14,99	1,41	3,75	1,41	25,96	5,19
IDP2	7,70	9,22	0,79	1,99	13,85	33,55	6,71
IDP3	2,71	12,94	2,81	4,03	12,58	35,06	7,01
Total	16,03	47,02	10,39	17,61	29,68		

4 Analysis of Results

To get an idea about the most touched and the least touched areas on the screen during learning and evaluation tasks, we divided the touched-screen into 16 areas (as in figure 5, r00...r03, r10...r13, r20...r23, r30...r33), and calculated the average of touches in each area for all users.

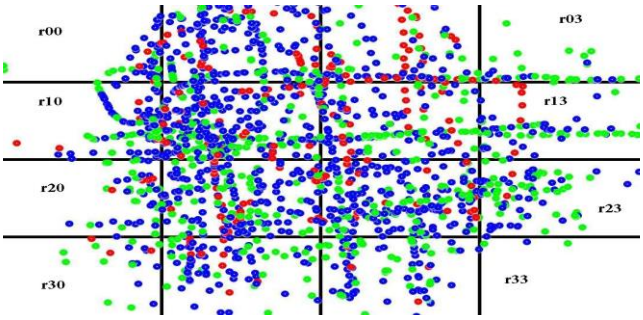


Figure 5. The 16 areas of the touched-screen of user with ID4 (Red points represent touched points with max pressure values, blue points represent touched points with pressure values less than the max and greater than the average, green points represent touched points with pressure values less than the average)

We have founded that the most touched areas are r12, r11, r22, r21, and the least touched areas are r30, r33, r32, r00 as described in figure 6. This information could be useful in next phases of our research in completing the graphical vibro-tactile language by putting the important information in the most touched areas.

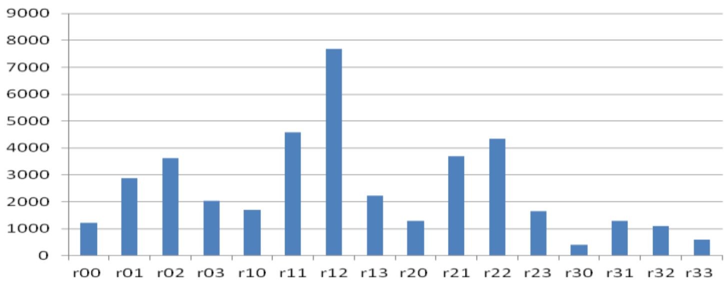


Figure 6. Most and least touched areas on the touched-screen

During analysis the results, we have noticed that there are a lot of differences between the pressure values for all users (Pressure value depends on the used tablet; in these experiments we have used Asus Model TF101 with Android operating system). To analyze pressure values, we calculated the max pressure value between all users, it was 3.19, and the average was 1.73, then we redrew the touched points for each user with considering that points with pressure values equal to the max value have been drawn in red color, points with pressure values less than the max and greater than the average have been drawn in blue color, and points with pressure values less than the average have been drawn in green color. Figure 5 represents an example of these points in different colors (for IDP3 evaluation task of the user who has id 4). After analyzing all the images drawn for all users, we have noticed that majority of red points (max pressure) are in images for which users gave right answers. This notice may be useful in designing our graphical vibro-tactile language, since we can observe when the user decreases or increases his touch pressure. The increasing of pressure may indicate that the user touches graphical elements interesting for him, and the decreasing may indicate that the user touches graphical elements non-interesting for him. During the tests we observed also that users try sometimes to scan the screen very quickly, it might be because they try to get a lot of information in a short time.

5 State of The Art for Segmenting Web Pages

Segmenting a web page is a fundamental phase for understanding its global structure. Extracting the global structure of web pages is useful in many domains such as information retrieval, data extraction, and similarity of web pages. Many approaches have been suggested for segmenting web pages, such as:

- DOM-based segmentation: it depends on analyzing the DOM tree (Document Object Model), and extracting the main structure of web pages depending on HTML tags as described in [14]. In their paper, the authors determine first the layout template of a web page, and then it divides the page into minimum blocks, and finally collects these minimum blocks into content blocks.

- Vision-based segmentation: this method divides the web page depending on the visual view of web page contents on a web browser. The well known tool VIPS (VIsion based Page Segmentation) [15] is based on this approach.

- Image processing based segmentation: this approach captures an image for the visual view of a web page, and then depends on image processing techniques to divide the captured image into sub-blocks [16] [17].

- Text-based Segmentation: this approach focuses on extracting only information about texts in a web page. After dividing the web page into blocks of texts, it could be possible to find the semantic relations between these textual blocks. This method is useful in many information retrieval domains such as question answering applications [18].

- Fixed-length segmentation: this approach divides the web pages into fixed length blocks (passages), after removing all HTML tags, where each passage contains a fixed number of words [19].

- Densitometric analysis based segmentation: this approach depends on methods applied in quantitative linguistics, where text-density refers to a measure for identifying important textual segments of a web page [20].

- Graph-based segmentation: This approach depends on transforming the visual segments of a web page into graph nodes, then applying many common graph methods on these nodes for combining them into blocks, or for making a clustering for these nodes. Some common works which depend on this approach are [21] [22].

6 Suggested Hybrid Segmentation Algorithm

Most of segmentation algorithms render first the web page using a web browser, and then segments the HTML elements into many blocks depending on the visual layout. Our constructed hybrid segmentation algorithm has been tested on 154 pages collected manually from many newspapers and e-commerce sites (www.leparisien.fr, www.lefigaro.fr, www.liberation.fr, www.amazon.fr, www.materiel.net, wwww.photobox.fr). The results have been integrated with our under-development Android program. The obtained results are promising because the segmentation algorithm can efficiently extract the web page blocks depending on the visual structure, and the algorithm can also convert correctly these blocks into zones (clustering the blocks). Our algorithm mixes three segmentation approaches, vision-based segmentation, DOM-based segmentation, and graph-based segmentation.

6.1 Vision-Based Approach

In this phase, we render the web page using Selenium [23] web driver and Mozilla FireFox browser, and we get its visual structure by Java-script code injection inside the HTML source code of the rendered web page. The obtained visual structure indicates the global hierarchy of the rendered web page. This phase assigns additional information for each DOM HTML element such as XPath and bounding box (location [X0, Y0], and size [height and width]).

The input of this phase is a web page HTML source code, and its output is an augmented HTML web page with injected information about bounding boxes and DOM XPath for each HTML element. In next sections, we refer to bounding boxes by blocks (i.e. each bounding box represents an HTML element, and may contain other bounding boxes).

6.2 DOM-Based approach

After obtaining the visual structure of a web page, we analyze its DOM structure by applying filters and re-organization rules for enhancing results of next phases. We divide the DOM elements depending on the specification of HTML5 content models proposed by the World Wide Web Consortium (W3C) [24]. This specification divides the HTML tags into 7 categories (Metadata content, Flow content, Sectioning content, Heading Content, Phrasing content, Embedded content, and Interactive content). The first applied filter is Metadata-Content-Filter, which deletes all the elements considered as metadata content elements except "title" tag. We delete the other tags because they do not contain useful visual information in next steps. The deleted tags are "base", "command", "link", "meta", "noscript",

“script”, and “style”. We do not delete the “title” node (the page title) because the textual information existing in this node could be useful for the user. We add also to this filtered group the following tags “comment”, “br”, and “doctype”.

The second applied filter is Dead-Nodes-Filter, where it deletes all HTML nodes that do not affect on the appearance, for example nodes with height or width equals to “0px” (zero pixel); or nodes with style properties ("display:none" or "visibility:hidden" or "hidden:true"). After applying the previous filters, we apply some re-organization rules in order to enhance visualizing the information in next phases. One example of these rules is Paragraph-Reorganization-Rule, where this rule re-constructs all paragraph child-nodes in one node contains the extracted sub-texts. We made this rule after analyzing many DOM structures, and observing that the text in some paragraph nodes is distributed between many child-nodes such as <i> (italic), <span>, <b> (bold), <strong>, <em> (emphasized), <small>, <mark> (marked), <del> (deleted), <ins> (inserted), <cite> (defining a title of work), <u> (underline), and <sub> (subscript). So extracting these sub-texts and collecting them in one text is useful and more efficient for visualizing them as one block in next phases rather than visualizing them as many separated blocks.

This rule was applied on all tags of type <P> (paragraph), and on all tags existed in the heading content category (specification of HTML5 content models), this group contains the following nodes (h1, h2, h3, h4, h5, h6, and hgroup). We also apply this rule on many other HTML tags which might contain textual child-nodes such as <a> (hyperlink tag), <abbr> (abbreviation), <acronym> (this tag is not supported in HTML5), <address> (contact information for the author/owner), <bdi> (Bi-directional Isolation), <button>, <label>, <li> (list element), and <q> (quotation). We used Jsoup tool [25] in order to access to a web page DOM structure, and getting its HTML hierarchy. The result of this phase is a filtered DOM-tree; each of its nodes is visible and contains XPath and bounding box information. The designed filters and re-organization rules were integrated with our framework, and then we applied these rules and filters on the vision-based segmented web pages (154 pages mentioned previously). After getting the filtered DOM-tree for each page, we represented the obtained bounding boxes on the used tablet Samsung GALAXY Tab 2 (10.1inch, dimensions HeightXWidthXDepth 175.3X256.7X9.7 mm, Android version 4.1.2), after making a scaling of sizes of bounding boxes to be appropriate with the new size of used tablet.

6.3 Graph-Based approach

After segmenting the web page depending on its visual structures and analyzing its DOM-structure, we apply a new graph-based segmentation algorithm called “Blocks2Zones Clustering” in order to group many similar blocks together in one zone. Clustering many blocks together is necessary in order to decrease the number of viewed blocks in some interfaces (instead of viewing many blocks, we view one zone which represents these blocks and then the user can navigate intra-elements inside the zone by double clicking on the graphical element of the chosen zone), and to group closed blocks in one zone (here, closeness depends on distances between blocks, this will be described next sections in details). The pseudo-code of the proposed algorithm is:

```
Blocks2Zones Clustering Algorithm
Input (Blocks, N° of desired Zones)
Output: Graph of N nodes (N Zones)
1- Transform the blocks into a graph (Un-Directed graph)
   1.1. Blocks → Nodes,
1.2. Make relations between the nodes, and assign weights for these relations.
2- If number of zones <= number of blocks
        end the algorithm,
  Else
  3- Find the node with the smallest size (node A)
     (size of node ==size of the rectangle bounds the node )
  4- For node A, find the connection which has the largest weight (node B).
  5- Group the nodes A, and B (A+B).
  6- Repeat steps 3-4-5 till number of blocks == number of zones
```

The output graph is described as following: G = (V, E), where G is undirected graph, V is a set of vertices (nodes or zones), and E is a set of edges (connections between zones), and |V|=n (number of desired zones). We define the set of vertices as $V=\{v_{i:}\ 1\leq i\leq n$ where n is the number of desired zones, and v_i is set of sub-zones}. We define the set of connections as following $E=\{e_j\colon e_j(v_{j1},v_{j2}) : v_{j1}\in V$, and $v_{j2}\in V\}$. To calculate weights between nodes, we used the Euclidian distance between the centers of two nodes (center of a node is the center of the rectangle which bounds the node). In following, an example of applying this algorithm on the main page of web site w3schools.com is given. In this example, we want that the algorithm segments the main page into 9 zones. The first iteration of the algorithm will divide the page into 13 zones (because there are 13 main nodes in the DOM structure – <Div> nodes) as illustrated in figure 7.a, and will construct a graph of 13 nodes as illustrated in figure 7.b.

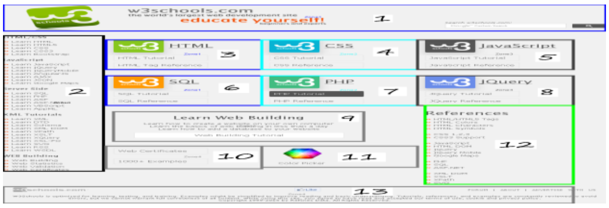


Figure 7.a Segmentation for 13 zones

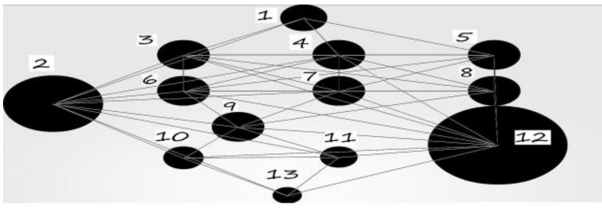


Figure 7.b Constructed graph

Figure 7. First segmentation of www.w3schools.com

Figures 8, and 9, illustrate next iterations of the algorithm and how the 13 zones are converted to 9 zones.

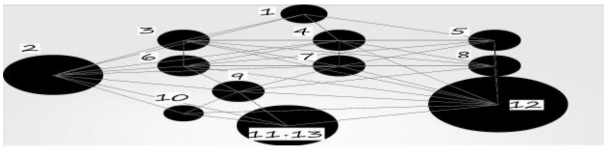


Figure 8.a 12 zones segmentation

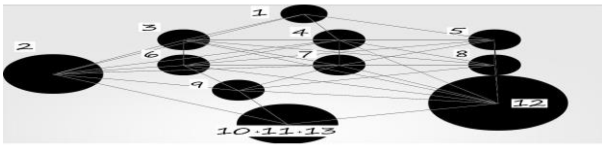


Figure 8.b 11 zones segmentation

Figure 8. 12, 11 zones segmentation of w3schools.com

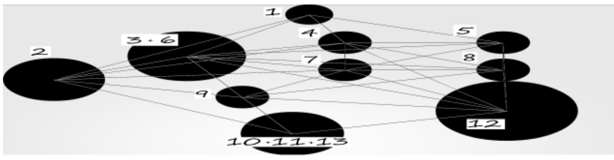


Figure 9.a 10 zones segmentation

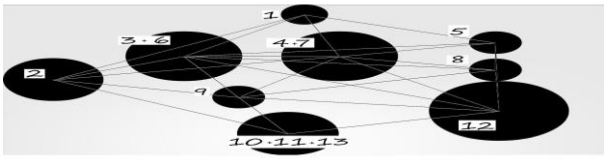


Figure 9.b 9 zones segmentation

Figure 9. 10, 9 zones segmentation of w3schools.com

Applying this hybrid segmentation algorithm on a filtered DOM-tree (obtained from applying Vision-based approach and DOM-based approach) converts a web page to a set of zones, each zone contains many other zones or blocks, since each block represents a visual structure of HTML element and may contain many other blocks. The purpose of the proposed vibro-tactile access protocol is then to transform semantics of symbols in these zones, or blocks, or HTML elements into vibrations with different frequencies and amplitudes.

7 Manual and Automatic Web Page Segmentation Differences

In order to evaluate our algorithm results, and to know how users understand web layout structures based on their visual perception, we made an experiment as following: we asked to 15 volunteers a manual segmentation of different kinds of web pages. The volunteers were of different ages (between 25 and 50 years old), and most of them were informatics specialists. We presented, for each volunteer, 4 printed copies (A4 size papers) of 8 web pages (2 pages from www.cdiscount.com, 2 pages from www.photobox.com, 2 pages from www.rueducommerce.fr, 1 page from www.w3schools.com, and 1 page from www.leparisien.fr). We asked each volunteer to segment the 4 copies of each web page into 3, 4, 5 and 6 zones with following considerations:

- all the pages are printed in gray scale in order to avoid affecting the colors on the segmentation process. We chose this option because the current version of our algorithm does not depend on color differences between blocks,
- users can segment the page using polygons with minimum of 4 points (triangles are not allowed),
- users can start segmenting from any part or direction of the page (left, right, top, down, or center),
- we asked users to write the ordering number of zones (inside or beside the zone) while they make the segmentation process; this is very useful for us to know how the users start segmenting the pages, and how they end it.

After collecting all manual segmented copies (480 papers : 15 users X 8 web pages X 4 copies), we noticed the following:

- 70% of users do not start the segmentation process for certain number of zones by the same way they start segmenting for other numbers of zones,

- 16% of papers have been segmented starting from the center of the page, 3.5% have been segmented starting from the bottom of the page, and the majority of papers 80.5% have been segmented starting from top of the pages,
- 20% of papers have been segmented vertically, and 80 have been segmented horizontally,
- 92.6% of segments are rectangles, and 7.4% of segments are polygons with more than 4 points,
- finally, we noticed that it is very difficult to detect a segmentation method common between all users; since each user segments the pages depending on his understanding of the web page layout structure, on his visual perception of the visible elements, and on his interests and visual experience.

We have run our algorithm with the 8 mentioned web pages. The algorithm segmented each web page into 3, 4, 5, and 6 zones. The comparison between manual and automatic segmentation is illustrated in tables 8 and 9.

Table 8. Strong-criterion based matching results of manual and automatic segmentation

	Page1	Page2	Page3	Page4	Page5	Page6	Page7	Page8	Total
3 Zones	1	0	0	6	3	0	9	1	20
4 Zones	1	5	4	3	3	0	3	2	21
5 Zones	4	4	1	4	3	0	0	2	18
6 Zones	4	1	2	2	4	1	0	1	15
Total	10	10	7	15	13	1	12	6	74

Table 8 illustrates matching results based on what we call “Strong criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if the results are 100% identical without any difference. We can conclude from table 1 that pages segmented automatically and manually into 4 zones are more matched than pages segmented into other numbers of zones. The percentage of identical matching depending on the strong criterion is 15.42% (74 identical matched results of 480 segmented copies).

Table 9 illustrates matching results based on what we call “Weak criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if at least 50% of the results are identical. We can conclude from table 2 that pages segmented automatically and manually into 3 zones are more matched than pages segmented into other numbers of zones. The matching percentage depending on the weak criterion is 47.5% (228 weak identical matched results of 480 segmented copies).

Table 9. Weak-criteria based matching results of manual and automatic segmentation

	Page1	Page2	Page3	Page4	Page5	Page6	Page7	Page8	Total
3 Zones	8	10	8	9	9	2	12	6	64
4 Zones	7	5	7	8	10	4	13	7	61
5 Zones	8	6	6	5	7	6	8	9	55
6 Zones	6	4	7	5	8	5	10	3	48
Total	29	25	28	27	34	17	43	25	228

8 Conclusion and Perspectives

In this paper, we summarized our current work which aims to design an approach for non-visual access to web pages on touch-screen devices. The designed vibro-tactile protocol transforms the information viewed on the screen and touched by users to vibration by transforming light contrasts of touched pixels into low-frequencies tactile vibrations. The obtained results are interesting, since we used in these experiments only one vibration motor of low level quality (phone vibrator), and the learning period was very short, so there are many enhancements to be achieved in next versions either on the hardware/software level or on the level of learning phase (increasing the number and quality of micro-vibrators, making more control on frequencies and amplitudes sent to micro-vibrators, adding vocal abilities to the current approach, …). Next steps in this research will be 1) adding some image processing techniques in order to enhance the proposed segmentation algorithm 2) adding advanced techniques in text summarization to facilitate navigating textual information, 3) making real experiments with blind VIP to study effects of suggested segmentation algorithm on their web navigation models.

References

[1] Maurel, F., Dias, G., Routoure, J-M., Vautier, M., Beust, P., Molina, M., Sann, C., 2012. “Haptic Perception of Document Structure for Visually Impaired People on Handled Devices”, Procedia Computer Science, Volume 14, 2012, Pages 319-329, ISSN : 1877-0509. DOI=http://dx.doi.org/10.1016/j.procs.2012.10.036

[2] Maurel, F., Vigouroux, N., Raynal, M., Oriola, B., 2003. “Contribution of the Transmodality Concept to Improve Web Accessibility”. In Assistive Technology Research Series, Volume 12, 2003, Pages 186-193. International

conference; 1st, Smart homes and health telematics; Independent living for persons with disabilities and elderly people. ISSN : 1383-813X.
[3] http://www.chromevox.com/ [Access 01/03/2015]
[4] http://www.synapseadaptive.com/gw/wineyes.htm [Access 01/03/2015]
[5] http://www.freedomscientific.com/ [Access 01/03/2015]
[6] https://play.google.com/store/apps/details?id=es.codefactory.android.app.ma.vocalizerfrfdemo&hl=fr [Access 01/03 /2014]
[7] https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=fr [Access 01/03/2014]
[8] http://www.apple.com/fr/accessibility/ [Access 01/03/2015]
[9] Alaeldin, A., Mustafa , Y., Sharief, B., 2012. "Tactile Web+avigator Device for Blind and Visually Impaired People". In Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies, Jordan, 2012. DOI=http://dx.doi.org/10.1109/AEECT.2011.6132519
[10] Rotard, M., Knödler, S., Ertl, T., 2005. "A Tactile Web Browser for the Visually Disabled". In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, 2005, pages 15-22. DOI= http://dx.doi.org/10.1145/1083356.1083361
[11] BouIssa , Y., Mojahid, M., Oriola, B., Vigouroux, N., 2009. "Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents". IEEE International Conference on Advances in Computational Tools for Engineering Applications, 2009, Pages: 591 – 594. DOI=http://dx.doi.org/10.1109/ACTEA.2009.5227855
[12] Lenay, C., Gapenne, O., Hanneton, S., Marque, C., Genouëlle, C., "Sensory Substitution, Limits and Perspectives. In Touch for Knowing Cognitive psychology of haptic manual perception", Amsterdam, 2003, Pages: 275-292.
[13] Tixier, M., Lenay, C., Le-Bihan, G., Gapenne, O., Aubert, D., "Designing Interactive Content with Blind Users for a Perceptual Supplementation System", TEI 2013, 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain, 2013, Pages 229-236. DOI= http://dx.doi.org/10.1145/2460625.2460663
[14] A. Sanoja, and S. Gançarski, "Block-o-Matic: a Web Page Segmentation Tool", BDA. Nantes, France. 2013. http://hal.archives-ouvertes.fr/hal-00881693/
[15] C. Deng, Y. Shipeng, W. Ji-Rong, and M. Wei-Ying, "VIPS: a Vision-based Page Segmentation Algorithm", Nov. 1, 2003, Technical Report MSR-TR-2003-79, Microsoft Research. 2003.
[16] D. Cai, X. He, W-Y. Ma, J-R. Wen, and H. Zhang, "Organizing WWW Images Based on the Analysis Of Page Layout And Web Link Structure", Microsoft Research Asia, Beijing, China, 2004.
[17] J. Cao, B. Mao, and J. Luo, "A segmentation method for web page analysis using shrinking and dividing", International Journal of Parallel, Emergent and Distributed Systems - Network and parallel computing, Volume 25 Issue 2, April 2010. Pages: 93-104, 2010. DOI=http://dx.doi.org/10.1080/17445760802429585
[18] N. Foucault, S. Rosset, and G. Adda, "Pré-segmentation de pages web et sélection de documents pertinent en Questions-Réponses", TALN-RÉCITAL 2013.
[19] J.P. Callan, "Passage- level Evidence in Document Retrieval", the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Pages: 302-310. Publisher: ISBN:0-387-19889-X, Springer-Verlag New York, Inc, 1994.
[20] C. Kohlschütter, and W. Nejdl, "A Densitometric Approach to Web Page Segmentation", USA, CIKM'08, Proceedings of the 17th ACM conference on Information and knowledge management, Pages: 1173-1182. 2008. DOI: http://dx.doi.org/10.1145/1458082.1458237
[21] D. Chakrabarti, R. Kumar, and K. Punera, "A graph-theoretic approach to webpage segmentation". Proceedings of the 17th international conference on World Wide Web, WWW'08, ACM, USA, Pages: 377-386. Publisher: ACM New York, NY, USA, 2008. DOI: http://dx.doi.org/10.1145/1367497.1367549
[22] X. Liu, H. Lin, and Y. Tian, "Segmenting Webpage with Gomory-Hu Tree Based Clustering", International Journal of Software Engineering and Its Applications, Vol 6, No 12, Pages: 2421-2425. 2011.
[23] http://www.seleniumhq.org/ [Access 01/03/2015]
[24] http://www.w3.org/TR/2011/WD-html5-20110525/content-models.html [Access 01/03/2015]
[25] http://jsoup.org/ [Access 01/03/2015]

An Empirical Study for Examining the Performance of Visually Impaired People in Recognizing Shapes through a Vibro-tactile Feedback

Waseem Safi Fabrice Maurel Jean-Marc Routoure Pierre Beust Gaël Dias
University of Caen Basse-Normandie – UNICAEN
14032 Caen- France
{waseem.safi, fabrice.maurel, jean-marc.routoure, pierre.beust, gael.dias}@unicaen.fr

ABSTRACT

In this paper, we present results of an empirical study for examining the performance of visually impaired people in recognizing shapes through a vibro-tactile feedback. The suggested vibro-tactile system maps different shades of grey to low-frequencies tactical vibrations. Performance data is reported, including number of errors, and qualitative understanding of the displayed shapes.

Categories and Subject Descriptors

[H.5.2] [**User Interfaces**] Graphical user interfaces (GUI), Haptic I/O, User-centered design, User interface management systems.

General Terms

Algorithms, Design, Human Factors, Languages.

Keywords

Visually impaired people, vibro-tactile feedback, low-frequencies tactical vibrations.

1. INTRODUCTION

An important accessibility drawback of current screen readers is the failure of visually impaired users to quickly get an overall sense of a webpage in terms of overall semantics, main message, structure, and interaction affordances [1]. Our work focuses on developing and evaluating a sensory substitution system based on a vibrotactile solution, cheap and efficient in noisy and public environments.

2. RELATED WORKS

Related works reviewed in this section are two-fold: first, we view previous researches on vibrotactile feedback technologies; second, we review some works leveraging the perception of layout and logical structures.

2.1. Vibrotactile Output

Many authors have proposed the attachment of vibrotactile actuators on users' body for working as mnemonic information [2]. Opticon is one of oldest systems that proposed a vibrotactile feedback [3]. Opticon translates the written word into a scanned display on the fingertips. Another interesting prototype designed for interactive tabletops has been proposed by [4], the prototype is represented by a device incorporates interactive haptics into tabletop interaction. A 2D tactile prototype has been suggested by [5] to train visually impaired people for their independent mobility. The prototype is associated to a 2D tactile array (vibration array) which consists of 16 vibrating elements arranged in a 4×4 manner. UbiBraille is a vibrotactile reading device has been designed by [6] that leverages the users' braille knowledge to read textual information. The main drawback of many proposed systems is that they need specific devices (such as solenoids and pin matrix), which are expensive and cannot be integrated easily to handled devices. Overall, previous researches have demonstrated that vibrotactile techniques could be useful in conveying semantic information, and in helping users to perform a variety of tasks [1]. Moreover, controlling vibration features such as amplitude, frequency, and duration, as well as the number of actuators allows production richer feedback. In our research, we explored that the usage of such features creating a mnemonic association with perception of layout and logical structures.

2.2. Perception Of Document Layout And Logical Structures

Perceiving the 2D structure of web pages greatly improves navigation efficiency and memorization as it allows high level reading strategies [1]. A tactile web browser for hypertext documents has been proposed by [7]. This browser renders texts and graphics for visually impaired people on a tactile graphics display and supports also a vocal feedback. Tactos is a perceptual interaction system [8], which consists of a tactile simulator, a graphics tablet with a stylus and a computer. A structured participatory-based approach has been proposed by [9], this approach aims to develop targeted haptic sensations for purposes of web page exploration. All these mentioned approaches aim to give visually impaired people the ability for perceiving logical structures. In this paper, we contribute a vibrotactile framework, which features (1) on hardware level: cheap, little intrusive, plugged-in easily with handled devices, modular for adding easily and synchronically many tactile electronic components - piezoelectric vibrators, solenoids matrices, peltiers- in order to get a large, rich and useful tactile semantics [10]; and (2) on software level: a graphical user interface, controlling wirelessly many parameters -vibration, amplitude, duration, rhythm, and repetition-, integrated easily with handled devices.

3. VIBROTACTILE FRAMEWORK

Our system "TactiNET" (Figure 1) provides a vibrotactile feedback when the blind user touches the tablet. To achieve the desired system, we have designed an electronic circuit, which controls a micro-vibrator placed anywhere on the body. A Bluetooth connection with an android tablet allows controlling the actuators. An Android dedicated program on the tablet views an image on the screen and detects information about the user touches (X, Y, Time, and Pressure). The intensity of the light emitted by the tablet at touched points is then transmitted to the embedded device in order to control the tactile stimuli.

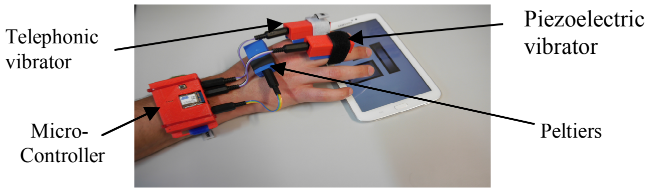


Figure 1: "TactiNET" prototype.

To validate our prototype and concepts of vibrotactile access to visual structures of web pages, we wanted first pretest our hypothesis: visually impaired people can explore and redraw simple grayscale shapes by using vibration motors.

4. PRE-TESTS PROTOCOL

We made experiments with 5 blind persons, each experiment consisted of: personal and technical questions, explanations of the test objective, a training task, and finally an evaluation task. The tested images (viewed on the tablet) contain some examples of expected results of web pages layout extraction, so success of distinguishing these shapes by blind users could be an indicator of their ability to distinguish web pages layouts. Figure 2 presents some images of evaluation phases.

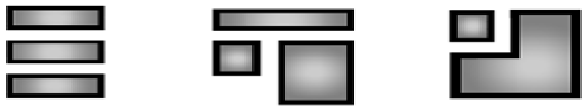


Figure 2: Images IDP1, IDP2, IDP3 (Evaluation Task).

Image IDP1 contains 3 rectangles with matched sizes and with vertical order, and the image IDP2 contains 3 rectangles with different sizes and many relations of directions, so testing these two images could indicate the ability of distinguishing sizes, and distinguishing relations of directions. Image IDP3 contains different shapes, so it could test the ability to distinguish different shapes in the same image. We created different transitions of gray scale for each shape to analyze the reaction through a continuous decreasing intensity from the micro-motor.

For each image in the evaluation task, we asked users to explore it and to answer for some questions (cf. in table 1, symbol ✓ represents a correct answer and the symbol X an incorrect one).

Table 1. Answers of questions for images IDP1, IDP2, IDP3

User-ID	ID0			ID1			ID2			ID3			ID4		
IDP	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Number shapes	✓	✓	X	✓	X	X	✓	X	✓	X	✓	✓	✓	✓	✓
sizes of shapes	X	X	X	X	X	X	✓	X	X	X	✓	✓	✓	✓	✓

At the end of the test, we asked each user to draw the recognized graphical patterns founded in each touched image (we used another tablet for redrawing). Some users could redraw forms nearly identical to original evaluation images. Figure 3 views the redrawing results of the user who answered successfully to all questions.

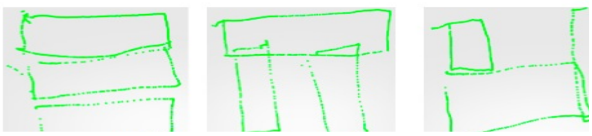


Figure 3. Results of redrawing images IDP1, IDP2, IDP3

5. ANALYSIS OF RESULTS

We notice from data in table 1 that the best performance is for the user with ID4, and this may be because this user is the youngest between others, and it could be because she was the only one who has already used touched devices (an iPhone, in her case working with VoiceOver), and this could be an indication that more training users more getting better results. Redrawing comparison with touched images can conclude: 1- ability of distinguishing sizes of shapes, because the degree of scaling between redrawn shapes is nearly equal to the degree of scaling between real shapes. 2- ability of distinguishing relations of directions (vertical order, left to, right to…), because relations of directions between redrawn shapes are nearly identical to relations of directions between real shapes.

6. CONCLUSION AND PERSPECTIVES

We have presented an embedded device dedicated for visually impaired people to explore simple grayscale shapes using vibrations. First results are promising: after a short training period, tested persons were able recognizing basic patterns. Many enhancements to be achieved such as increasing the number and quality of micro-vibrators, and more controlling frequencies and amplitudes.

7. REFERENCES

[1] Maurel, F., Dias, G., Routoure, J-M., Vautier, M., Beust, P., Molina, M., Sann, C. "Haptic Perception of Document Structure for Visually Impaired People on Handled Devices", Procedia Computer Science, Volume 14, Pages 319-329, 2012.

[2] Kammoun, S., Jouffrais, C., Guerreiro, T., Nicolau, H., Jorge, J., "Guiding Blind People with Haptic Feedback". In Pervasive Workshop on Frontiers in Accessibility for Pervasive Computing, New Castle, UK, 2012.

[3] Bliss, J., Katcher M., Rogers C., Shepard R., "Optical-to-Tactile Image Conversion for the Blind", Published in: Man-Machine Systems, IEEE Transactions, Volume:11, Issue:1, March; pages: 58 – 65, 1970.

[4] Marquardt, N., Nacenta, M., Young, J., Carpendale, S., Greenberg, S., Sharlin, E., "The Haptic Tabletop Puck: tactile feedback for interactive tabletops." Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces. ACM, Pages 85-92, 2009.

[5] Dakopoulos, D., Bourbakis, N., "Towards a 2D tactile vocabulary for navigation of blind and visually impaired.", Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics, Pages: 45-51, ISSN: , 1062-922X, 2009.

[6] Nicolau, H., Guerreiro, J., Guerreiro, T., Carriço, L., "UbiBraille: Designing and Evaluating a Vibrotactile Braille-Reading Device", Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, Article No. 23. ACM New York, USA 2013.

[7] Rotard, M., Knödler, S., Ertl, T., "A Tactile Web Browser for the Visually Disabled". In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, pages 15-22, 2005.

[8] Tixier, M., Lenay, C., Le-Bihan, G., Gapenne, O., Aubert, D., "Designing Interactive Content with Blind Users for a Perceptual Supplementation System", TEI 2013, 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Pages 229-236, Barcelona, Spain, 2013.

[9] Kuber, R., Yu, W., McAllister, G., "Towards developing assistive haptic feedback for visually impaired Internet users." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Pages 1525-1534, 2007.

[10] Brewster, S., Brown, L., "Tactons: structured tactile messages for non-visual information display". In Proceedings of the fifth conference on Australasian user interface-Volume 28, Pages: 15–23. Australian Computer Society, Inc., 2004.

An Empirical Study for Examining the Performance of Visually Impaired People in Discriminating Ranges of Frequencies through Vibro-tactile Feedbacks

Waseem Safi Fabrice Maurel Jean-Marc Routoure Pierre Beust Michèle Molina Coralie Sann
University of Caen Normandie – UNICAEN
14032 Caen- France

{waseem.safi, fabrice.maurel, jean-marc.routoure, pierre.beust, michele.molina, coralie.sann}@unicaen.fr

ABSTRACT
In this paper, we present results of an empirical study for examining the performance of blind individuals in discriminating ranges of frequencies through vibro-tactile feedbacks. The suggested vibro-tactile system maps different shades of grey to one pattern low-frequencies tactical vibrations.

Author Keywords
Visually impaired people; vibro-tactile feedback; low-frequencies tactical vibrations.

ACM Classification Keywords
[H.5.2] [User Interfaces] Graphical user interfaces (GUI), Haptic I/O, User-centered design, User interface management systems.

INTRODUCTION
An important accessibility drawback of current screen readers is the failure of individuals who are blind or visually-impaired to quickly get an overall sense of a web page in terms of overall semantics, main message, structure, and interaction affordances [1]. Our work focuses on developing and evaluating a sensory substitution system based on a vibro-tactile solution, cheap and efficient in noisy and public environments.

RELATED WORKS
Related works reviewed in this section are two-fold: first, we view previous researches on vibro-tactile feedback technologies; second, we review some works leveraging the perception of layout and logical structures.

Vibro-tactile Output
Many authors have proposed the attachment of vibro-tactile actuators on users' bodies for working as mnemonic information [2]. Opticon is one of oldest systems that proposed a vibro-tactile feedback [3]. Opticon translates the written word into a scanned display on the fingertips. Another interesting prototype designed for interactive tabletops has been proposed by [4], the prototype is represented by a device incorporates interactive haptics into tabletop interaction.

A 2D tactile prototype has been suggested by [5] to train blind people for their independent mobility. The prototype is associated to a 2D tactile array (vibration array) which consists of 16 vibrating elements arranged in a 4×4 manner. UbiBraille is a vibro-tactile reading device [6] that leverages the users' braille knowledge to read textual information. The main drawback of many proposed systems is that they need specific devices which cannot be integrated easily to nowadays handled devices.

Perception of Document Layout and Logical Structures
Perceiving the 2D structure of web pages greatly improves navigation efficiency and memorization as it allows high level reading strategies [1]. A tactile web browser for hypertext documents has been proposed by [7]. This browser renders texts and graphics for visually impaired people on a tactile graphics display and supports also a vocal feedback. Tactos is a perceptual interaction system [8], which consists of a tactile simulator, a graphics tablet with a stylus and a computer. All these mentioned approaches aim to give visually impaired people the ability for perceiving logical structures.

VIBROTACTILE FRAMEWORK
Our system "TactiNET" (Figure 1) provides one pattern vibro-tactile feedback when the blind user touches a tablet. To achieve the desired system, we have designed an electronic circuit, which controls many micro-vibrators placed anywhere on the body. A Bluetooth connection with an android tablet allows controlling the actuators. An Android dedicated program on the tablet views an image on

the screen and detects information about the user's touches (X, Y, Time, and Pressure). The intensity of the light emitted by the tablet at touched points is then transmitted to the embedded device in order to control the tactile stimuli.

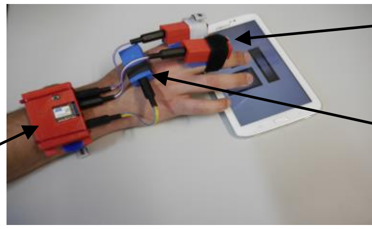


Figure 1: "TactiNET" prototype.

A series of experiments validated our prototype and concepts of vibro-tactile access to visual structures of web pages [9]. First pre-tests validated our hypothesis: visually impaired people can explore and redraw simple grayscale shapes by using vibration motors [9]. The series of experiments described in this paper aims to select a range of frequencies most perceptible by sighted and visually impaired persons. These ranges of frequencies will be used in generating vibro-tactile feedbacks that represent contrasts of visual elements in web pages.

PRE-TESTS PROTOCOL

38 sighted children (average 8.26 years) and 25 adults participated in the study. The 25 adult participants were composed of 20 sighted persons (average 29.8 years) and of 5 blind persons (average 57 years). Each participant had to navigate on the touch-screen, split in two equal-sized parts. They were asked whether the vibration feedbacks generated when touching the first part of the screen were identical to the vibration feedbacks generated when touching the second part of the screen. To run the experiments, two tablets of type Samsung GALAXY Tab 2 (10.1 inch) have been used. The first tablet, connected with the prototype device of TactiNET, is dedicated to the haptic exploration of the participant. The second tablet is dedicated to the experimenter to generate the various pattern of vibration send to the first tablet. The two tablets are connected by a Bluetooth connection.

Participants navigate on the tablet by using the index of their preferred hand (left or right). The actuator to perceive the vibrations are placed on the non-preferred hand. The experimenter asked them a single question: « are the vibration feedbacks generated when navigating the left part equal to that generated when navigating the right part?"

Five frequencies have been chosen to be evaluated as reference frequencies: 101.5625Hz; 203.125Hz; 304.6875Hz; 406.25Hz; and 500Hz. Each value of these five reference frequencies have been evaluated with two conditions of amplitude variability V0 and V5. Variability V0 means that the amplitude value is always 255 for all the vibration feedbacks generated when the participant navigates any part of the tablet screen. Variability V5 means that the amplitude value is between 255 and 250 (255-5). When the variability V5 is activated, a random integer value between 0 and 5 will be generated for each touch on any part of the tablet screen. This random integer value will be subtracted from the maximum amplitude value 255. The objective of adding these two types of variability is evaluating the framework sensitivity in public or noisy environments. Using the framework in public environments or noisy situations (trains, buses, walking situations, etc.) might affect on the performance of the users. A simple change in the amplitude value could be a simulation of generating some noisy factors.

After selecting the reference frequency values, the values of non-reference frequencies to be compared with the reference values have been determined. Each reference frequency value is compared with values of 10 series (5 ascendants and 5 descendants) of non-reference frequencies. Each series consists of 13 successive values. The difference between every two successive values in the same series is 7.8125Hz. The reference frequency value is the center value of each series. The experimenter starts always with an ascendant series.

ANALYSIS OF RESULTS

The differential perceptual thresholds for each reference frequency have been calculated under two conditions: variability V0, and variability V5. The reference frequency 304.6875Hz has the smallest differential perceptual thresholds either with the variability V0 or with the variability V5.

An ANOVA statistical analysis has been calculated taking into account two types of variabilities, and the visual status of the participants (sighted or blind). The statistical analysis indicated an effect for changing the reference frequency: $F(4,28) = 3.58$, $p = 0.017$, $\alpha = 0.81$. An analysis post-hoc with the test of Bonferroni has indicated that the differential perceptual threshold for the reference value 304.6875 Hz is the least important comparing with other differential perceptual thresholds. This result means that "the range of frequencies that are close to value 304.6875 Hz is more perceptible and discriminated than ranges of frequencies close to other tested reference frequencies".

An ANOVA analysis has been conducted to estimate the effect of the amplitude variability (V0 or V5) on the differential perceptual thresholds. The analysis did not indicate any effect of the variability: $F(1,8) = 0.052$, $p = 0.83$, $\alpha=0.05$. This means that the simple amplitude variabilities do not affect on the performance of the participants.

CONCLUSION AND PERSPECTIVES

We have presented an embedded device dedicated for visually impaired people to explore simple grayscale shapes using vibrations. First results are promising and close to other studies which confirm that navigating the digital devices is more sensible to vibrations with frequencies starting from 200-250 Hz. Many enhancements to be achieved such as increasing the number and quality of micro-vibrators, and applying the obtained results on the designed framework.

REFERENCES

[1] Maurel, F., Dias, G., Routoure, J-M., Vautier, M., Beust, P., Molina, M., Sann, C. "Haptic Perception of Document Structure for Visually Impaired People on Handled Devices", Procedia Computer Science, Volume 14, Pages 319-329, 2012.

[2] Kammoun, S., Jouffrais, C., Guerreiro, T., Nicolau, H., Jorge, J., "Guiding Blind People with Haptic Feedback". In Pervasive Workshop on Frontiers in Accessibility for Pervasive Computing, New Castle, UK, 2012.

[3] Bliss, J., Katcher M., Rogers C., Shepard R., "Optical-to-Tactile Image Conversion for the Blind", Published in: Man-Machine Systems, IEEE Transactions, Volume:11, Issue:1, March; pages: 58 – 65, 1970.

[4] Marquardt, N., Nacenta, M., Young, J., Carpendale, S., Greenberg, S., Sharlin, E., "The Haptic Tabletop Puck: tactile feedback for interactive tabletops." Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces. ACM, Pages 85-92, 2009.

[5] Dakopoulos, D., Bourbakis, N., "Towards a 2D tactile vocabulary for navigation of blind and visually impaired.", In proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics, Pages: 45-51, ISSN: , 1062-922X, 2009.

[6] Nicolau, H., Guerreiro, J., Guerreiro, T., Carriço, L., "UbiBraille: Designing and Evaluating a Vibrotactile Braille-Reading Device", Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, Article No. 23. ACM New York, USA 2013.

[7] Rotard, M., Knödler, S., Ertl, T., "A Tactile Web Browser for the Visually Disabled". In Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA, pages 15-22, 2005.

[8] Tixier, M., Lenay, C., Le-Bihan, G., Gapenne, O., Aubert, D., "Designing Interactive Content with Blind Users for a Perceptual Supplementation System", TEI 2013, 2013, in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Pages 229-236, Barcelona, Spain, 2013.

[9] Safi, W, Maurel, F., Routoure, J-M, Beust, P, Dias, G., 2015. "An Empirical Study for Examining the Performance of Visually Impaired People in Recognizing Shapes through a Vibro-tactile Feedback", ASSETS'15, October 26-28, 2015, Lisbon, Portugal.