

2F4T: Audio and Force Feedback for the Non-Visual Exploration of a Work of Art

2F4T : Audio et retour de force pour l'exploration non visuelle d'une œuvre d'art

LILIA DJOUSSOUF, LITIS, University of Rouen Normandie, Normandy University, France

THOMAS BERTHELIN LE TELLIER, GREYC, University of Caen Normandie, Normandy University, France

KATERINE ROMEO, LITIS, University of Rouen Normandie, Normandy University, France

CHRISTELE LECOMTE, LITIS, University of Rouen Normandie, Normandy University, France

GAËL DIAS, GREYC, University of Caen Normandie, Normandy University, France

DHIA EDDINE MERZOUGUI, GREYC, University of Caen Normandie, Normandy University, France

FABRICE MAUREL, GREYC, University of Caen Normandie, Normandy University, France



This work presents a multimodal approach aimed at improving the accessibility of two-dimensional works of art for blind or partially sighted individuals. By coupling the F2T haptic device, based on force feedback, with the TagThunder system, which enables spatialized audio exploration of segmented web pages, users can autonomously interact with visual content. This combination supports intuitive hierarchical navigation by integrating tactile guidance, free exploration, and voice feedback. The goal is to enrich users' sensory and cognitive experience without the need for a companion.

Ce travail présente une approche multimodale visant à améliorer l'accessibilité des œuvres d'art bidimensionnelles pour les personnes aveugles ou partiellement aveugles. En couplant le dispositif haptique *F2T*, basé sur le retour de force, au système TagThunder, qui permet une exploration audio spatialisée de pages web segmentées, l'utilisateur peut interagir de manière autonome avec des contenus visuels. Ce couplage favorise une navigation hiérarchique intuitive, combinant guidage tactile, exploration libre et retour vocal. L'objectif est d'enrichir l'expérience sensorielle et cognitive des utilisateurs, sans nécessiter la présence d'un accompagnateur.

CCS Concepts: • **Human-centered computing** → **Haptic devices**; *Auditory feedback*; **Accessibility technologies**; **Accessibility systems and tools**; • **Computing methodologies** → **Natural language processing**.

Additional Key Words and Phrases: Non Visual Multimodal Interaction, Force Feedback, Spatialized Audio Description

Authors' Contact Information: Lilia DJOUSSOUF, lilia.djoussouf@univ-rouen.fr, LITIS, University of Rouen Normandie, Normandy University, Rouen, France; THOMAS BERTHELIN LE TELLIER, Thomas.berthelin-letellier@etu.unicaen.fr, GREYC, University of Caen Normandie, Normandy University, Caen, France; KATERINE ROMEO, Katerine.Romeo@univ-rouen.fr, LITIS, University of Rouen Normandie, Normandy University, Rouen, France; CHRISTELE LECOMTE, Christele.Lecomte@univ-rouen.fr, LITIS, University of Rouen Normandie, Normandy University, Rouen, France; GAËL DIAS, gael.dias@unicaen.fr, GREYC, University of Caen Normandie, Normandy University, Caen, France; DHIA EDDINE MERZOUGUI, dhia-eddine.merzougui@unicaen.fr, GREYC, University of Caen Normandie, Normandy University, Caen, France; FABRICE MAUREL, Fabrice.Maurel@unicaen.fr, GREYC, University of Caen Normandie, Normandy University, Caen, France.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

Mots Clés et Phrases Supplémentaires: Interaction multimodale non visuelle, Retour de Force, Description Audio spatialisée

ACM Reference Format:

Lilia DJOUSSOUF, THOMAS BERTHELIN LE TELLIER, KATERINE ROMEO, CHRISTELE LECOMTE, GAËL DIAS, DHIA EDDINE MERZOUGUI, and FABRICE MAUREL. 2025. 2F4T: Audio and Force Feedback for the Non-Visual Exploration of a Work of Art. 1, 1 (July 2025), 5 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 Introduction

Les visiteurs de musées qu'ils soient aveugles ou partiellement aveugles sont souvent limités dans leur expérience face aux œuvres d'art. Les moyens proposés par les musées, quand ils existent, se centrent sur un nombre limité de représentations accessibles. Les audioguides lorsqu'ils sont mis à disposition sont activés par visée avec un QR code ou une carte RFID posée à proximité visuelle des œuvres. Leur utilisation nécessite pour une majorité de personnes la présence d'un accompagnateur. C'est notamment le cas pour des œuvres d'art qui ne peuvent pas directement être touchées ou manipulées par soucis de préservation. Tandis que les sculptures peuvent faire l'objet de réplcation, il est plus difficile de représenter des objets d'art bidimensionnels tels que des peintures, des tapisseries, des broderies etc. Ainsi, dans l'optique de permettre une expérience plus autonome sur des œuvres d'art bidimensionnelles, diverses technologies ont été mises en œuvre pour favoriser leur interactivité. L'objectif de la démonstration présentée ci-contre est de coupler deux dispositifs innovants pour améliorer l'expérience sensorielle d'immersion des personnes aveugles ou partiellement aveugles, dans la découverte d'œuvres d'art. En premier lieu, la plateforme TagThunder [4], développée par le laboratoire GREYC, intègre dans sa dernière version [7] différents modules visant l'accès aux pages Web à partir de retours audio spatialisés produits selon différentes stratégies interactives d'exploration (*skimming* et *scanning* non visuels). D'autre part, cet article présente le système *F2T* (Force Feedback Tablet) [3] conçu par le laboratoire du LITIS, qui se base sur le retour de force pour faciliter le guidage des utilisateurs aveugles ou partiellement aveugles. Suite à l'étude utilisateur sur le dispositif de la *F2T*, il a été mis en avant que la méthode de guidage passive a aidé les participants à comprendre l'œuvre. En couplant la plateforme TagThunder au système de la *F2T* on obtient l'outil 2F4T (Force Feedback Tablet et TagThunder Tree), les auteurs cherchent à présenter une méthode d'exploration autonome, complète, intuitive et interactive. Dans cet article, nous allons présenter le fonctionnement de ces deux outils, ensuite la méthode de couplage sera explicitée.

2 Système d'interaction multimodale

2.1 Le dispositif haptique F2T

La *F2T* (Figure 1a) est un dispositif se basant sur la mécanique des imprimantes 3D. Elle est composée d'un joystick, sur lequel l'utilisateur doit poser son index, relié à deux moteurs permettant un mouvement selon les axes x et y . Le logiciel de la *F2T* traite des images virtuelles qui sont chargées selon des scripts écrits au préalable [3]. L'utilisateur a accès à deux méthodes principales d'exploration, l'exploration guidée et l'exploration sans guidage [2].

L'**exploration guidée** se base sur une perception passive des mouvements en jouant notamment sur la proprioception de l'utilisateur. Elle se divise, en deux niveaux : l'exploration guidée globale (Figure 1b) et l'exploration guidée détaillée (figure 1c), pour faire le lien avec l'idée de « gist » tactile [8], première impression [6]. Le premier niveau de « gist » tactile introduit par l'exploration guidée globale présente les éléments sémantiquement importants de l'image en présentant des chemins globaux pour décrire grossièrement ces formes et leur position dans l'image. Le deuxième niveau de « gist » tactile se base sur la recherche puis la sélection d'un élément, à l'aide d'attracteurs placés sur des objets

dans l'image donnant accès à des contours précis. Ces chemins plus précis correspondent aux contours de l'élément. Ils sont à chaque fois accompagnés par des retours donnant les descriptions verbales de ce qui est désigné. En effet, dans l'expérience d'une œuvre, les retours haptiques nécessitent d'avoir des descriptions audios [9].

L'exploration sans guidage, quant à elle, implique une perception active de l'utilisateur. Il peut librement explorer l'espace de travail dans le but d'accéder à des informations en déplaçant le joystick. Pour cette phase, des images virtuelles sont chargées d'appliquer divers effets de frictions, ou de flux, qui ont pour but d'améliorer la sensation d'immersion en ayant des indices sur des éléments dynamiques présents dans l'image (Figure 1d). Le logiciel de la *F2T* repose essentiellement sur un système d'événements permettant le chargement dynamique d'objets spécifiques, tels que des images virtuelles pour la génération d'effets haptiques ou des fichiers de trajectoires définissant les contours d'objets. La présentation de ces méthodes d'exploration hiérarchique est principalement linéaire (exploration guidée globale -> exploration guidée détaillée -> exploration sans guidage). Ainsi, il nous semblerait intéressant de mieux prendre en compte la volonté de l'utilisateur illustrée par ces mouvements et d'exploiter les fonctionnalités de *TagThunder*.

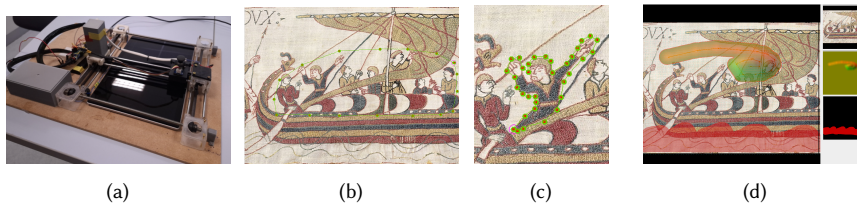


Fig. 1. Présentation du système *F2T*. (1a) Image du dispositif *F2T* sous sa version avec des courroies (1b) Exemple de trajectoire (en vert) pour l'exploration guidée globale et (1c) guidée détaillée. (1d) Interface de la *F2T* avec une image issue de la Tapisserie de Bayeux avec un effet de flux dans la voile et un effet de friction fluide pour les vagues

2.2 La plateforme TagThunder

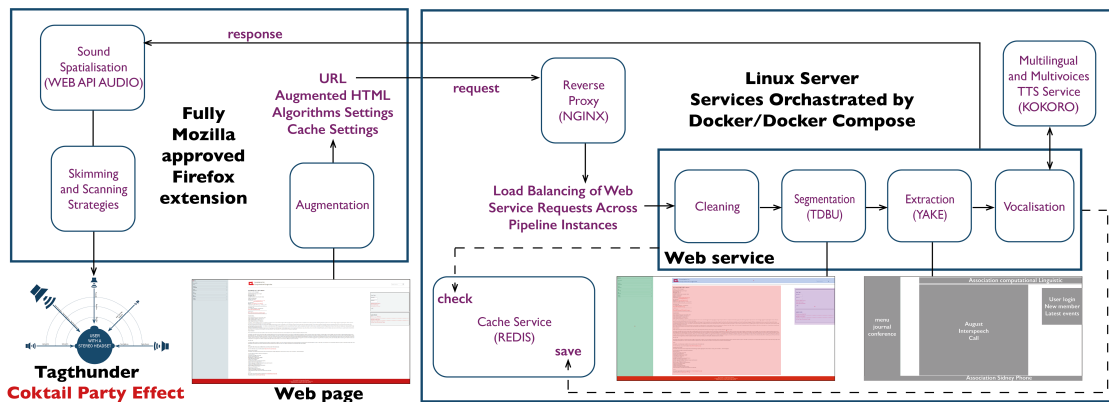


Fig. 2. Architecture serveur/plugin firefox de la plateforme TagThunder

Métaphore de l'effet cocktail party : le cadre expérimental TagThunder repose sur une analogie étendant le concept psychologique de l'« effet cocktail party ». Il désigne la capacité à concentrer son attention auditive sur un flux verbal

au sein de l'environnement bruyant d'un rassemblement. Nous développons cette métaphore en considérant la relation entre un utilisateur non-voyant et les différentes zones d'une page web, analogue à la relation entre un invité situé dans une pièce et divers groupes de discussion. Ainsi, chaque zone est perçue comme un groupe de discussion potentiel, dont le sujet exprimé vocalement correspond à l'ensemble des mots-clés pertinents de cette zone. Le paysage sonore construit est ainsi composé de plusieurs séquences de termes clés répétées simultanément selon différentes sources audio concurrentes et spatialisées. Il s'agira alors pour l'utilisateur de piocher dans ce "tonnerre de mots" (sorte de pendant oral du classique "nuage de mot") l'information nécessaire pour se diriger de manière autonome vers une thématique repérée et donc, par construction, vers une zone d'intérêt de la page Web. Notre hypothèse principale est que, sous la condition d'une couche interactive appropriée, la déambulation au sein de cet environnement sonore permettra à terme la découverte et l'ancrage de stratégies de compréhension rapide et non visuelle de l'organisation logico-thématique des pages Web.

Chaîne de traitement : *TagThunder* est conçu comme une chaîne de traitement, illustrée dans la Figure 2. Premièrement, la page web subit une phase de prétraitement afin de générer un document HTML enrichi contenant des informations visuelles, logiques et de contenu, lequel est ensuite nettoyé pour éliminer les éléments structurels non pertinents. Deuxièmement, l'algorithme TDBU [10] segmente la page Web afin d'identifier des zones pertinentes. Troisièmement, YAKE! [1] extrait les termes-clés de chaque zone en se basant sur leur importance lexicale et leur mise en forme locale. Enfin, les termes sont vocalisés simultanément à l'aide du modèle de synthèse vocale Kokoro¹ et spatialisés en 3D via l'API Web Audio² selon la disposition des zones, comme proposé dans [5].

Plate-forme d'évaluation : du point de vue de l'utilisateur, les processus d'interaction avec la page Web développées dans *TagThunder* ont pour objectif d'évaluer différentes stratégies non visuelles de survol (*skimming*) et de recherche rapide (*scanning*) d'informations (Cf. Annexe technique). La structure calculée par le système, sous la forme d'une hiérarchie de zones d'intérêt (*TagThunder Tree*), autorise une exploration arborescente reproduisant l'idée d'un zoom/dézoom non visuel. La chaîne de traitement peut alors être relancée au niveau de granularité souhaité lorsque l'utilisateur sélectionne une zone d'intérêt.

3 Adaptation du système TagThunder avec la F2T

Le système 2F4T, dans sa composition, permettra d'une part de rechercher de manière autodéterminée grâce à la plate-forme *TagThunder* des zones de l'oeuvre qui attirent l'attention de l'utilisateur, et d'autre part, d'explorer grâce au système *F2T* de façon guidée et multimodale les détails pertinents. L'intégration initiale de la chaîne de traitement de *TagThunder* avec le dispositif *F2T* permet l'envoi de commandes vers son logiciel, dans le but de déclencher une exploration hiérarchique à partir d'une interface web contenant une image, où chaque élément visuel est enrichi d'une description sémantique. Une étape préalable de segmentation est réalisée afin d'extraire les entités visuelles pertinentes à décrire (Fig. 3a). Chaque élément identifié est ensuite associé à une description spécifique, permettant une navigation structurée et informative. Un exemple de segmentation, accompagné des descriptions textuelles correspondantes, est illustré en Figure 3b et 3c. La démonstration se déroule sur une page web contenant des images d'oeuvres d'art, que l'utilisateur, quelle que soit sa capacité visuelle, peut sélectionner et explorer à l'aide de la plate-forme *TagThunder*. Lorsqu'une zone de l'oeuvre attire son attention, il peut effectuer une commande pour déclencher le parcours des contours sur la tablette *F2T*.

¹<https://huggingface.co/hexgrad/Kokoro-82M>

²https://developer.mozilla.org/fr/docs/Web/API/Web_Audio_API

L'utilisateur peut également sélectionner une zone spécifique afin de la segmenter et de l'explorer plus en profondeur. Chaque sous-zone pourra ensuite être explorée individuellement à l'aide de la tablette F2T.

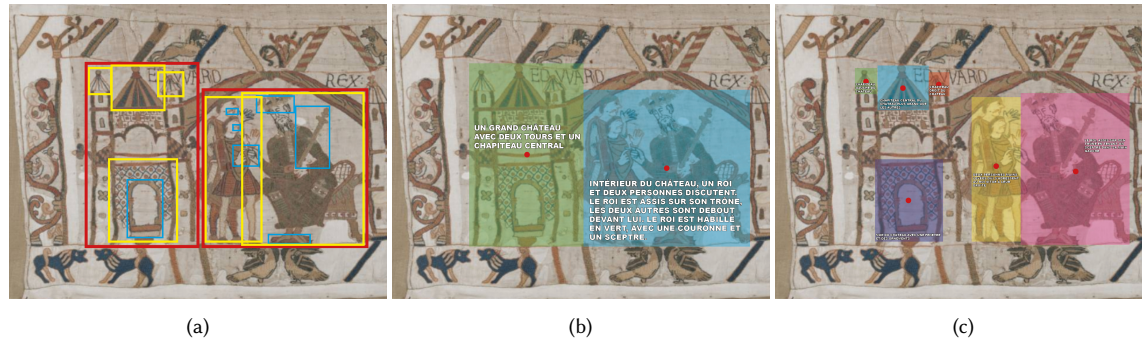


Fig. 3. Traitement d'une image de la Tapisserie de Bayeux en page web segmentée en zones (3a), partitionnées et associées à leur description textuelle (3b) sur plusieurs niveau de granularité (3c).

4 Remerciements

Cette étude est financée dans le cadre d'un stage M2 NormaStic³. Elle a pour objectif de poursuivre le projet ANR-IMG (ANR-20-CE38-0007)⁴ qui s'intéresse à l'accessibilité pour tous, des oeuvres d'art dans les musées.

References

- [1] Ricardo Campos, Vitor Mangaravite, Arian Pasquali, Alipio Jorge, Célia Nunes, and Adam Jatowt. 2020. YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences* (2020), 257–289.
- [2] Lilia Djoussouf, Nicolas Charafe, Ilies Mohammed-Riad Zizi, Christele Lecomte, and Katerine Romeo. 2024. Hierarchical exploration of artworks through 2-DOF haptic device: a usability study. In *Eurohaptics (Session WIP)*. doi:10.5281/zenodo.12601235
- [3] Simon L. Gay, Edwige Pissaloux, Katerine Romeo, and Ngoc-Tan Truong. 2021. F2T: A Novel Force-Feedback Haptic Architecture Delivering 2D Data to Visually Impaired People. *IEEE Access* 9 (2021), 94901–94911. doi:10.1109/ACCESS.2021.3091441 Conference Name: IEEE Access.
- [4] Jean-Marc Lecarpentier, Elena Manishina, Maxence Busson, Fabrice Maurel, and Stephane Ferrari. 2016. Tag Thunder : plateforme de démonstration et d'expérimentation (Tag Thunder : demonstration and experimentation platform). In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. volume 5 : Démonstrations*, Laurence Danlos and Thierry Hamon (Eds.). Paris, France, 52–54. <https://aclanthology.org/2016.jeptalnrecital-demo.19/>
- [5] Fabrice Maurel, Gaël Dias, Stéphane Ferrari, Judith-Jeyafreeda Andrew, and Emmanuel Giguët. 2019. Concurrent Speech Synthesis to Improve Document First Glance for the Blind. In *2nd International Workshop on Human-Document Interaction associated to ICDAR*.
- [6] Fabrice Maurel, Gaël Dias, Waseem Safi, Jean-Marc Routoure, and Pierre Beust. 2020. Layout Transposition for Non-Visual Navigation of Web Pages by Tactile Feedback on Mobile Devices. *Micromachines* 11, 4 (April 2020), 376. doi:10.3390/mi11040376
- [7] Dhia Eddine Merzougui, Nilesh Tete, Fabrice Maurel, Gaël Dias, Mohammed Hasanuzzaman, Aurélien Bournonville, Edgar Madelaine, Thomas Berthelin, Francois Ledoyen, Laure Lejeune-Poutrain, François Rioult, and Jeremie Pantin. 2025. Concurrent Speech and Auditory Tag Clouds for Non-Visual Web Interaction. In *26th edition of the Interspeech Conference*. Rotterdam, Netherlands. <https://hal.science/hal-05110274>
- [8] Edwige Pissaloux, Lilia Djoussouf, Katerine Romeo, Ramiro Velazquez, Simon Gay, Ngoc-Tan Truong, and Son Dao. 2022. A new approach to assist Virtual Image accessibility for Visually Impaired People. In *9th International Conference on human computer interaction. HCI 2022. Virtual (On-line)*, Sweden. <https://hal.science/hal-03665582>
- [9] Katerine Romeo, Hannah Thompson, and Marion Chottin. 2022. Inclusive Multimodal Discovery of Cultural Heritage: Listen and Touch. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Georgios Kouroupetroglou, Katerina Mavrou, Roberto Manduchi, Mario Covarrubias Rodriguez, and Petr Penáz (Eds.). Springer International Publishing, Cham, 278–285. doi:10.1007/978-3-031-08648-9_32
- [10] Waseem Safi. 2019. *Non-visual Vibro-tactile Navigation of Web Pages on Touch Screen Devices*. PhD Thesis. University of Caen Normandie (France).

Received ; revised ; revised

³<https://www.normastic.fr/>

⁴<https://anr.fr/Projet-ANR-20-CE38-0007>